2002

Spread spectrum-based video watermarking algorithms for copyright protection

Serdean, Cristian Vasile

SPREAD SPECTRUM-BASED VIDEO WATERMARKING ALGORITHMS FOR COPYRIGHT PROTECTION

by

CRISTIAN VASILE SERDEAN

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

Satellite Research Centre

Department of Communication and Electronic Engineering

Faculty of Technology, University of Plymouth

In collaboration with BBC Research & Development

November, 2002

Abstract

SPREAD SPECTRUM-BASED VIDEO WATERMARKING ALGORITHMS FOR COPYRIGHT PROTECTION

by

Cristian Vasile Serdean

Digital technologies know an unprecedented expansion in the last years. The consumer can now benefit from hardware and software which was considered state-of-the-art several years ago. The advantages offered by the digital technologies are major but the same digital technology opens the door for unlimited piracy. Copying an analogue VCR tape was certainly possible and relatively easy, in spite of various forms of protection, but due to the analogue environment, the subsequent copies had an inherent loss in quality. This was a natural way of limiting the multiple copying of a video material. With digital technology, this barrier disappears, being possible to make as many copies as desired, without any loss in quality whatsoever. Digital watermarking is one of the best available tools for fighting this threat.

The aim of the present work was to develop a digital watermarking system compliant with the recommendations drawn by the EBU, for video broadcast monitoring. Since the watermark can be inserted in either spatial domain or transform domain, this aspect was investigated and led to the conclusion that wavelet transform is one of the best solutions available. Since watermarking is not an easy task, especially considering the robustness under various attacks several techniques were employed in order to increase the capacity/robustness of the system: spread-spectrum and modulation techniques to cast the watermark, powerful error correction to protect the mark, human visual models to insert a robust mark and to ensure its invisibility. The combination of these methods led to a major improvement, but yet the system wasn't robust to several important geometrical attacks. In order to achieve this last milestone, the system uses two distinct watermarks: a spatial domain reference watermark and the main watermark embedded in the wavelet domain. By using this reference watermark and techniques specific to image registration, the system is able to determine the parameters of the attack and revert it. Once the attack was reverted, the main watermark is recovered. The final result is a high capacity, blind DWT-based video watermarking system, robust to a wide range of attacks.

Contents

List of figures

List of tables

Glossary

A/D. Analogue/Digital.

ACC. Accumulator.

AWGN. Additive White Gaussian Noise.

BER. Bit Error Rate.

BCH. Bose - Chaudhuri - Hocquenghem.

bpf. Bits Per Frame.

bps. Bits Per Second.

BPSK. Binary Phase Shift Keying.

CPTWG. Copy Protection Technical Working Group.

CRC. Cyclic Redundancy Code.

CWT. Complex Wavelet Transform.

D/A. Digital/Analogue.

DCT. Discrete Cosine Transform.

DFT. Discrete Fourier Transform.

DINT. De-interleaver.

DSSS. Direct Sequence Spread Spectrum.

DV. Digital Video.

DWT. Discrete Wavelet Transform.

EBU. European Broadcasting Union.

ECG. Electrocardiography.

EEG. Electroencephalography.

EKG. Electrocardiography.

ENC. Encoder.

EZW. Embedded Zero-tree Wavelet.

FEC. Forward Error Correction.

FFT. Fast Fourier Transform.

FMT. Fourier Mellin Transform.

HH, HL. High High, High Low.

HPF, hpf. High Pass Filter.

HVS. Human Visual System.

IDWT. Inverse DWT.

IDCT. Inverse DCT.

IFFT. Inverse FFT.

i.i.d. Independent Identically Distributed.

INT, INTER. Interleaver.

INT^{-1}. De-interleaver.

ITU. International Telecommunication Union.

JND. Just Noticeable Difference.

JPEG. Joint Photographic Experts Group.

LH, LL. Low High, Low Low.

LLT. Log-Log Transform.

lpf. Low Pass Filter.

LPT. Log-Polar Transform.

LSB. Least Significant Bit.

MJPEG. Motion JPEG.

MPEG. Moving Picture Experts Group.

MRS. Magnetic Resonance Spectra.

MTF. Modulation Transfer Function.

NC. Number of Cross-correlations.

PAL. Phase Alternate Line.

PCCC. Parallel Concatenated Convolutional Code.

pdf. Probability Density Function.

PN. Pseudo Noise.

POMF. Phase Only Matched Filter.

QPSK. Quadrature Phase Shift Keying.

RS. Reed-Solomon.

RSC. Recursive Systematic Code.

RST. Rotation, Scale and Translation.

SCCC. Serial Concatenated Convolutional Code.

SDMI. Secure Digital Music Initiative.

SISO. Soft Input Soft Output.

SNR. Signal to Noise Ratio.

SPIHT. Set Partitioning In Hierarchical Trees.

SPOMF. Symmetrical POMF.

STFT. Short Time Fourier Transform.

TC. Turbo Codes.

TEL. Tolerable Error Level.

VCR. Video Cassette Recorder.

VDP. Video Dependent, Perceptual.

VHS. Video Home System.

VIP. Video Independent, Perceptual.

VWG. Video Watermarking Group.

WM. Watermark.

X-corr. Cross-correlation.

ACKNOWLEDGMENTS

I would like to express my sincere thanks to my supervisors, Professor Martin Tomlinson and Dr. Graham Wade for their support and guidance during my PhD. Additionally, I am very grateful to Dr. Graham Wade for his feedback, and for providing me with his expertise in writing scientific publications and in particular for my PhD thesis.

I would like to address special thanks to my colleague and friend Dr. Adrian Ambroze for all his help during my PhD, and particularly for supplying me with some of his C code. I gratefully acknowledge all his programming help and advice.

I gratefully acknowledge the support of EPSRC, which through a research studentship made all this work possible. I would like to extend my gratitude to BBC R&D and TRL Technology Ltd. who contributed financially to my studentship.

Many thanks to my friends: Zaki Ahmed, Adrian Ambroze, Peter van Eetvelt, Keith Flack, Bogdan Ghita, Licha Mued, Phil Rodwell, James Slader and Levente Toth for their company, help and "considerable amount of good quality humour that compensated for the weather".

I would also like to extend my gratitude to all the administrative staff from University of Plymouth, particularly to Mrs. Brenda Garraghan, Mrs. Barbara Fuller and Ms. Sue Locke for all their help and efficiency in solving many bureaucratic issues.

Many thanks to all my Romanian Professors for providing me with good technical education, in spite of the difficult economic conditions. Special thanks to Prof. Monica Borda for offering me the opportunity of coming to Plymouth as an MSc exchange student, and for all her subsequent help.

Finally but not at least, I would like to express my warmest thanks to my family, for all their support and sacrifice during all my undergraduate and graduate studies. My endless gratitude goes to my mother who encourages and supports me, and yet misses me day by day; and to my late father, God rest his soul, for all his effort and invaluable teachings.

I would also like to thank to my relatives for their support, particularly to Fam. Raicu and Neag.

DECLARATION

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

This work was financed by the EPSRC grant GR/M37691 and additional industrial contributions from BBC Research & Development, Kingswood Warren and TRL Technology Ltd., Tewkesbury.

This research programme included an extensive literature survey and attendance of relevant international conferences. The work has been regularly presented at the research seminars organised by the University of Plymouth, in several international conferences and in technical meetings with BBC, TRL Technology and other partners.

This work was concretised in publishing two journal papers (one as a co-author) and six international conference papers (one as a co-author). The full list of publications is provided in section 8.3.

Signed: .....

Date: 09 DEC 2002

This thesis is dedicated to my mother Victoria and to the loving memory of my father Vasile (1940-1990).

Chapter 1

"Do what you feel in your heart to be right – for you'll be criticized anyway. You'll be damned if you do, and damned if you don't."

Eleanor Roosevelt (1884-1962)

Introduction

Nowadays virtually all multimedia production and distribution is digital. The advantages of digital media, for creation, processing and distribution are all well known: superior quality, more quicker and easier to edit and modify, possibility of software processing rather than the more expensive hardware alternative (if the real time processing is not a requirement), and maybe the most important advantage is the unlimited copying of digital data without any loss of quality whatsoever. This latter advantage is not desired at all by the media producers and content providers, in fact is perceived like a major threat, because it may cause them considerable financial loss.

Once the digital technology is widely available to the public, the piracy suddenly becomes a major issue. This generates the need for protecting the copyrighted material against piracy. Some typical examples are the recent court battles between the music industry and Napster, Kazaa and Morpheus. The movie and music industry are particularly keen to develop any system which will stop users copying the digital media. Especially after the introduction of Internet sharing technologies which allow users from the entire planet to share any kind of digital media between them (like Napster, Gnutella, Morpheus and others) the record labels are trying to stop this trend by virtually any method possible. The cheapest and most effective ways in the long term are the non-technical methods like endless threats and law suites and using their huge influence to promote harsher copyright protection laws. When everything else fails, the only remaining alternative are the technical methods in the form of various copyright protection techniques.

The perfect example is the case of the VCR (Video Cassette Recorder). Probably not too many people know that when the VCR was marketed, the record labels tried to stop the technology by filling a law suit, on the grounds that the VCR technology could be used to copy protected material. Fortunately for us and for the technical development which led to technologies like CD-R and DVD-R they didn't succeed, but they managed instead to impose legal taxes on the blank recording media (VCR tapes, CD-R's, CD-RW's and others), taxes which are included in the final price paid by the user. Although this system is not currently implemented in UK, it is in force in most of the European countries and USA. Since in this instance the legal way failed, the technical approach was the only alternative left. The result was the development of the Macrovision copy-protection system which proved to be quite efficient against the casual VCR piracy.

Even if the user in fact pays for the right of copying digital materials, the record labels recently introduced a copyright protection system for the audio CD's which actually tries to stop the users from copying their legitimate CD's and even playing the CD's on a computer. Actually, this protection system developed by the Israeli company Midbar, deliberately introduces during the fabrication process a substantial number of errors on the disk, in fact so many, that even the powerful error correction capability of the computer drives is defeated. This is a rather "sad" method which destroys the very core of the digital technology, lowering not only the quality, but also the reliability of the disk. In fact the legitimate buyers were so upset, that the record labels had to withdraw the disks from the market, and as a result Philips who holds the rights for the CD-ROM standard won't allow the record labels to use the CD logo on this kind of protected CD's. These methods are rather obtrusive and have the "quality" of angering the legitimate customers, and even more, they are apparently illegal in those countries in which the customers are paying levies on the recording media.

Unlike these "crude" methods, digital watermarking is an unobtrusive way of protecting such material and for audio, images and video it operates by hiding a perceptually invisible signal into the host signal.

1.1 Historical Roots

The roots of watermarking as an information hiding technique can be traced in the ancient Greece as *Steganographia* or *steganography* as we know it now. The origin of the word steganography comes from the Greek στεγανός – "steganos" γραφειν – "graphein" which

literally means "covered writing". Many dictionaries are not even mentioning the word and few of them which are including the word are wrongly explaining it as cryptography. In fact although these two notions are related they are quite different. While cryptography focuses on encrypting a message so it can be read only by its intended recipient, steganography, on the other hand, keeps the message secret by hiding the fact that the message exists at all.

The historical roots of steganography and the beginnings of watermarking are well described in the literature [Singh, 1999], [Kobayashi, 1997], [Swanson et al, 1998-1], [Hartung et al, 1999-2], [Wolfgang et al, 1997 and 1999], [Langelaar et al, 2000]. Maybe one of the most well known example of a steganographic technique which is still widely used even today are the omnipresent watermarks which could be found in virtually any bank note, in different official documents and even in some stamps.

From the perspective of copyright protection, most of the researchers are making a clear distinction between steganography and watermarking from the robustness point of view. One important property of watermarking which is not characteristic to the steganography is the robustness to attacks, by attacks understanding virtually any technique which tries to modify/alter/remove/destroy the watermark.

The first efficient analogue copyright protection system appeared in the early eighties, soon after the VCR made its public debut. The system called Macrovision, basically adds some "parasite" pulses to the video signal during blanking periods, in such a manner that the TV sets are not affected. These pulses are instead seriously perturbing the sync circuitry within any modern recorder with an analogue input, making impossible to copy a protected tape. Although the system is not too difficult to defeat, it has been proven to be an excellent tool against casual piracy. As a result, the Macromedia copyright protection was embedded into the DVD standard in order to protect the analogue outputs of the DVD players.

The debut of the digital watermarking techniques was made in the early nineties [Tanaka et al, 1990]. The actual term "watermarking" was introduced in [Tirkel et al, 1993]. Although the word has its roots in maritime terminology, a more appropriate translation is that of transparent or invisible marking. The term "watermarking" survived in spite of few other alternatives like: labelling, stamping and tattooing.

The watermarking passed more or less unnoticed, until in the mid nineties when the digital distribution of media content started to emerge on a larger scale and the content providers together with the copyright owners started to become very interested in copyright protection technology in order to reduce or stop the major piracy threat involved by digital

media distribution. The year 1995-1996, marked a real boom in watermarking research, as **Table 1-1** shows.

1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
2	0	2	1	8	21	59	106	224	259	258	258

Table 1-1 The number of publications on digital watermarking during the years according to [Peter Meerwald, 2002].

1.2 Classification of the Watermarks

It is not an easy job to classify the watermarks and the watermarking techniques. They could be classified from so many perspectives that an exhaustive classification is almost out of question. The classification presented here will be limited to the cases relevant to this thesis.

Classifying the watermarking techniques function of the media to be marked is maybe the first one worth to be mentioned. There can be text, audio, image and video watermarking. Some attempts were done to watermark the software as well. It has to be mentioned from the beginning that this thesis is mainly dealing with video watermarking and occasionally with image watermarking, in order to compare it with other schemes.

To see exactly where we stand, **Figure 1-1** presents the main classes of watermarks and highlights the case presented during this thesis. So from now on, the word "watermark" refers to the robust and invisible case.

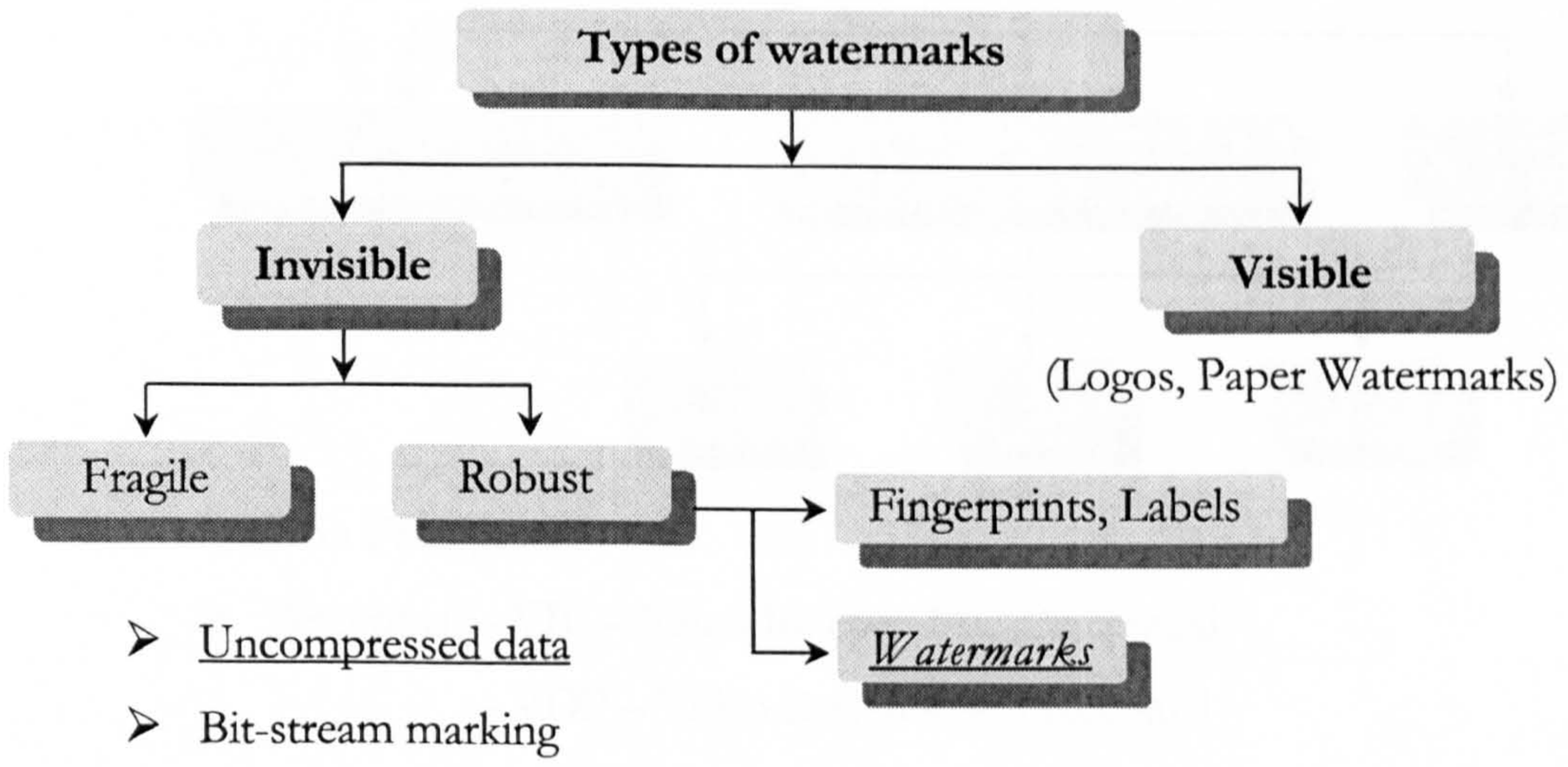


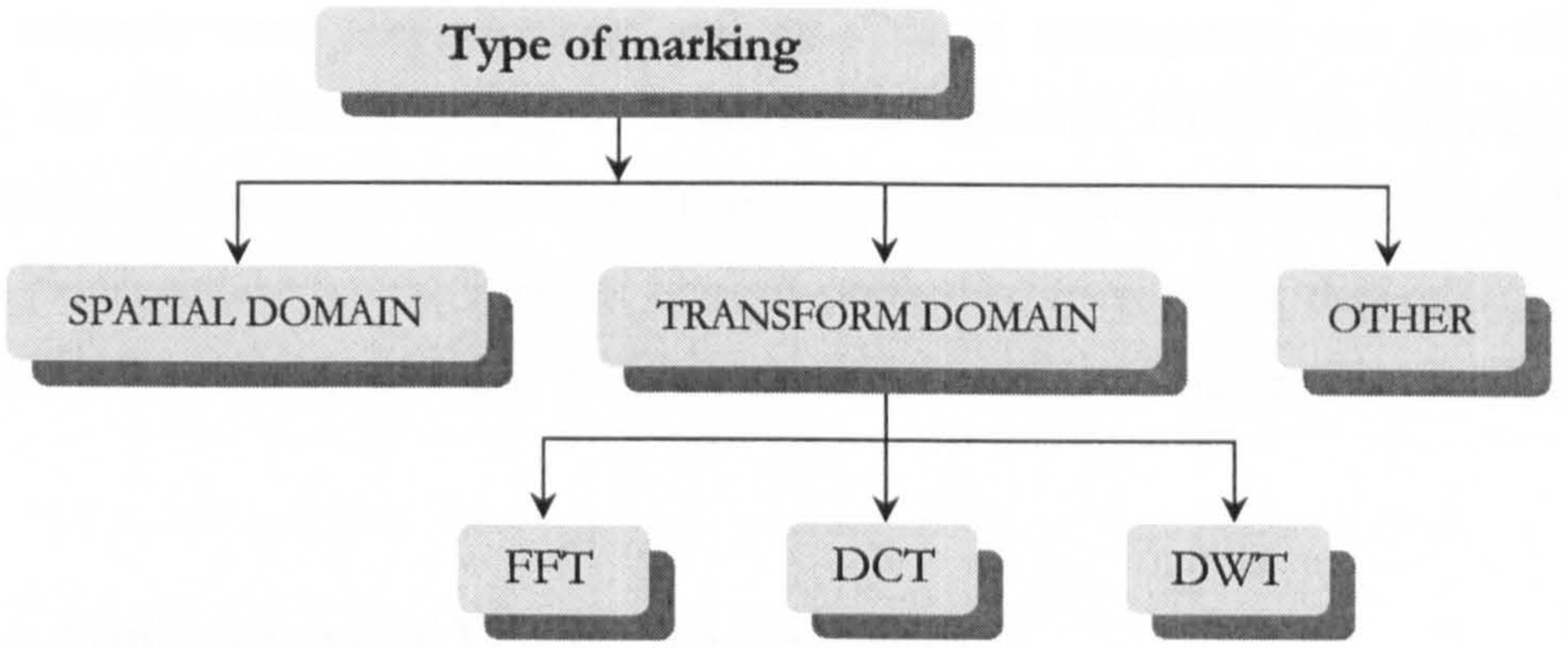
Figure 1-1 Classification of copyright protection marking

As **Figure 1-1** shows, there are two types of watermarks: the visible ones, like different logos either on paper or on a TV screen and the most important one, the invisible or transparent watermarks, which cannot be perceived by the human sensory system. An invisible watermark can be either robust or fragile. The use of a fragile watermark is important when one wants to verify if the protected media was tampered with or not. This type of watermark is especially designed to be as fragile as possible, so even the slightest modification of the marked media will destroy it, indicating that someone tampered with the media in question. This type of watermark is like a CRC (cyclic redundancy code).

The main class of watermarks – the robust ones – can be classified function of the purpose of the watermark, and therefore this could be an indication of the length of the watermark. Usually the fingerprints and labels are in fact serial numbers, typically quite short (e.g. 64 bits), which will uniquely identify the marked media (of course they could carry additional information if desired). If the embedded mark is more than a label, or if is quite long or has other purposes, then is specified by the general term: "watermark".

Additional to this classification, in the specific case of video watermarking, depending of where exactly the watermark is embedded during the distribution chain, we could have a watermark embedded in the uncompressed data or in the MPEG2 bit stream (without decoding and re-encoding the video). The case described during this thesis is the watermarking of the RAW uncompressed data.

Figure 1-2 shows another useful classification of the watermarking techniques,



- Non Perceptual
- Perceptual – VIP – Video Independent, Perceptual
  – VDP – Video Dependent, Perceptual

Figure 1-2 Classification of digital watermarking techniques function of the domain where the watermark embedding is performed.

function of the domain where the watermarking is embedded. The first attempts to watermark an image/video sequence were done in the spatial domain. This is quite simple, quick and obviously has DSP implementation advantages, but suffers from the lack of a good visual model. Generally speaking, embedding a watermark in the transform domain is more attractive because of the higher degree of freedom, and because it is naturally suited for perceptual marking based upon the *Human Visual System* (HVS). Visual models were specifically developed for this domain, mainly in the context of JPEG and MPEG2 compression. The usual transforms considered are the DFT/FFT (Discrete Fourier Transform/Fast Fourier Transform), DCT (Discrete Cosine Transform) and DWT (Discrete Wavelet Transform). The DFT/FFT can be a very attractive choice, offering the possibility of phase modulation and some useful invariance properties. Its main disadvantage is that in order to obtain a real image it is necessary to maintain complex conjugate symmetry, which effectively halves the potential marking capacity. Another disadvantage is the lack of good HVS models. Probably the most popular choice is the DCT transform, and this may remain so especially when watermarking MPEG2 compressed video, since the DCT is the basis of MPEG2. A more advanced choice is DWT. The wavelet transform is structurally very close of the way in which is thought the HVS is working, being by itself a HVS model, so the need of a complex HVS model is more reduced. The DWT marking is a lot more flexible than both FFT and DCT and has major advantages compared with these two, as will be shown in Chapter 6.

As **Figure 1-2** suggests, another classification of watermarking techniques could be with respect to the perceptual algorithm used at the embedding of the watermark. The first watermark schemes were non-perceptual and as a direct consequence not too robust. Embedding the watermark according to a perceptual algorithm brings the advantage of reduced visibility and increased robustness. Depending of the HVS model used, this could be image/video independent for the simplest ones or preferably image/video dependent in the case of a more advanced model, which therefore can adapt itself to the particular image/video.

1.3 The Applications of Digital Watermarking

As a general definition, digital watermarking can be regarded as hiding a message signal into a host signal without any perceptual distortion of the host signal. In theory, watermarking should protect a file permanently, like an invisible tattoo which cannot be removed without significantly damaging the protected data. This makes it the ideal tool for copyright protection,

file tracking and monitoring. Therefore most of the research was carried out with respect to these applications. Taking a closer look at the range of possible applications, we can identify:

- ✔ *Copyright protection:* To protect its intellectual rights, the owner of the digital media uses a watermark which carries copyright information in order to be able to prove later on (in a court of law) that a third party infringed its copyrights. In other words, the embedded watermark is used to show ownership of the digital media.

- ✔ *Fingerprinting:* A fingerprint can be compared with a serial number, and makes it easy to trace the source of illegal copies. The owner embeds a different fingerprint in each copy of the media sold to a different customer. In this way is very easy to establish who have broken the licence agreement by supplying the media to a third party and therefore to detect unauthorised duplication.

- ✔ *Broadcast monitoring:* By watermarking the broadcasted media prior to transmission, one can continuously monitor all the broadcasting channels, and identify all the intellectual property violations. This application is of major interest to broadcast corporations, who are very keen to protect their broadcasted material, especially since some of it has a huge market value, making it particularly prone to piracy. One eloquent example is the News, where even a photo or a short amateur video sequence can worth a huge amount of money. A slightly different application could be an automatic registration and monitoring of broadcasted radio and TV programs such that the appropriate royalties are automatically paid to the right owners.

- ✔ *Copy protection and usage control:* This is another major application of watermarking particularly wanted by the record labels. The information contained in the watermark directly controls the digital recording devices and can allow or prohibit the recording/viewing or even the storage of the media. The watermark could indicate for example "never copy" or "copy allowed". DVD access control could be one of the candidates who will benefit from this technology. By inserting a dynamic watermark which includes a number indicating how many copies of the media can be allowed, and by decreasing that number every time when the media was copied, one can restrict the copying of the media to a certain predefined number. This could be useful in electronic media distribution.

- ✔ *Authentication and integrity verification:* Using a fragile watermark, one can establish if the watermarked media is the original or if someone tampered with it. In this way is possible to authenticate the media or to validate its integrity.

Although the main interest at the moment is in the copyright protection area, mostly due to the strong involvement of the industry and record labels, there could be many more applications suitable for digital watermarking, most of them being less demanding in terms of constraints compared with the typical copyright protection applications:

- *Hidden content labelling and indexing:* A watermark could be used to provide subsidiary information about an image or a video, such as who is the subject, the date of production and any additional comments. This information could be used for searching and indexing purposes, making a lot easier to find and organise digital media.

- *Medical safety:* Closely related with the previous, this application could be a very useful safety measure, by allowing insertion of patient specific data into the medical images. By its nature, this application requires that the watermark should alter as little as possible the original data, or in other words a very light marking.

- *Media enhancements:* The watermarking could be used for adding different enhancements to the existing media. An example is the adding of various information to audio and video files, like details about the singer, a biography of the actors, subtitles and virtually anything else. All this "extras" are "free", in the sense that they could be applied to any existing format while keeping the backward compatibility and they do not require any additional storing space, and therefore the dimension of the media is preserved.

- *Data hiding and secret communications:* Watermarking techniques can be certainly used for transmission of secret, private messages exactly like the steganography was used during the course of the history for non digital media. With governments trying to restrict the encryption techniques, sooner or later this will be the ideal tool for secret communications. This is probably the worse nightmare of the secret services and various other governmental agencies who want to be able to intercept and control everything, since watermarking techniques are certainly not making this easier for them.

1.4 Requirements of a Video Watermarking System

The requirements for a watermarking system are obviously different for different applications, but even so, referring to the copyright protection context, it is possible to identify several common requirements for a watermarking system:

✔ *Perceptual transparency:* One of the most important requirements in most applications is to embed a watermark in a perceptually transparent manner. This means that one cannot make a difference between the marked and the original media under typical viewing conditions.

✔ *Robustness to attacks:* The aim of any watermarking system is to embed the mark in such a way that it cannot be removed, unless the media will be severely degraded during this process. Therefore the watermark should be able to withstand a wide range of attacks some of them unintentional, but most of them intentional, e.g. designed to modify/alter/remove/destroy the watermark. This is usually one major requirement in many watermarking applications, with the exception of fragile watermarking, which requires exactly the opposite: the watermark has to be as fragile as possible.

✔ *Blind recovery:* Due to the dimensions of video materials, especially in the uncompressed case presented in this thesis, video watermarking techniques specifically require blind or oblivious recovery of the mark. This means that the original cannot be used in the recovery process, since due to the huge dimensions involved it is impossible (or at least extremely expensive) to maintain a database with all the originals. This requirement is paramount.

✔ *Bit rate of data embedding algorithm (data payload of the watermark):* Depending on the application, this requirement can range anywhere between 1 data bit (this type of scheme is only able to tell if the image was marked or not with a specific watermark) and few hundreds or more data bits. Closely related with the bit rate of the embedding algorithm is watermark granularity, which represents the minimum segment of data containing a unit of watermark. For example, the typical requirement for a broadcast monitoring system is at lest 64 data bits per a video segment of 1 second (25 frames for PAL).

✔ *Security:* The embedding procedure must be secure or in other words, an unauthorised user should not be able to detect and remove the watermark. Most of the schemes are embedding the watermark according to a secret key which controls the insertion of data in the host signal. This respects the Kerckhoff's principle (1883) from cryptography which states that the security should lie in a secret key rather than in the algorithm's secrecy. Even if one is familiar with the scheme but doesn't know the secret key, an unauthorised detection/removal of the watermark should be impossible in a reasonable amount of time.

✔ *Copyright protection and ownership deadlock:* When the watermark is used to establish ownership of the media, the scheme should be able to resolve the rightful ownership even when multiple ownership claims are made. This is still an open problem, as today no scheme

can unambiguously determine the ownership of a marked media if it does not use the original media or another copy in the detection process, which is evidently not the case in video watermarking due to the blind recovery requirement. The problem was first described in [Craver et al, 1997]. The deadlock arises when a pirate simply adds his watermark to an already marked media and then claims the ownership. The problem is to establish (in the absence of the original) who watermarked the media first.

All these requirements are related to each other and quite contradictory. Probably visibility versus robustness is the most common example. A very robust watermark assumes a more heavily marked video, which obviously leads to increased visibility. Therefore the necessity of a trade-off is obvious. The other requirements are weakening even more a watermarking system. The blind recovery and the necessity of embedding many bits in a limited minimum segment are clear examples of that. Security for example is closely related with robustness: if the watermark is not secure, the system is not robust at all.

Several other application dependent requirements can be added to those already mentioned here: the algorithm should work in real time (hardware) and should be as simple and efficient as possible, preferably cheap to implement in hardware and easy to interface with the existing electronic devices and of course a reliable detection of the watermark. Depending of the application, there could be additional economical and technical requirements.

1.5 Structure of the Thesis

The thesis is organised as follows:

Chapter 1 offers a basic introduction in digital watermarking and its applications. The historical roots of the watermarking are traced back in time and the watermarks are then classified, emphasising the case discussed in this thesis. An overview of the possible applications of digital watermarking, together with the major players in this field is also provided. Finally, the main requirements of a video watermarking system are presented and briefly analysed.

Chapter 2 is an overview of the existing image/video watermarking methods. From spatial domain to wavelet-based watermarking, this is the place where the most important methods and algorithms relevant to the thesis are underlined. The basis and starting hypotheses of the

thesis are set here, emphasising the reasons for choosing one alternative rather than another. The reader is then introduced into the specific problem of video watermarking. This chapter also provides a specific introduction to the particularities of attacks in the video watermarking context, in close relationship with the EBU's recommendations. The chapter ends by highlighting the contributions of the thesis.

Chapter 3 presents in detail the structure and the algorithm of a spatial domain spread spectrum video watermarking system. This chapter establishes the foundations of the spread spectrum watermarking system. The main structure and many of the components described here are used for building the DCT scheme from Chapter 5. Starting with the basic uniform marking, the system is gradually improved by embedding the watermark in a more efficient way (e.g. HVS dependent) and by using a sliding correlator for watermark recovery. The effects of pre-filtering and various block sizes on the performance of a spatial domain watermarking scheme are also investigated.

Chapter 4 presents some basic communication principles related to channel capacity and forward error correction (FEC) codes. The Shannon's channel capacity theorem and its practical implications are also discussed, as well as the ways of getting closer to this limit. It is shown that in order to achieve better performance in a communication system, e.g. to get closer to the Shannon's limit, one can use the new state-of-the-art FEC codes (e.g. Turbo Codes). A short introduction to Turbo codes and their characteristics and performance is also provided in this chapter. The watermarking is seen by the information theory perspective and therefore by applying this theory and by using Turbo coding, the performance of the watermarking system is greatly improved (as Chapter 5 and Chapter 6 will show).

Chapter 5 discusses the case of watermarking in the DCT domain, together with several methods of increasing the capacity/robustness of the system. To achieve this goal, the system uses both advanced HVS models for watermark embedding and state-of-the-art FEC (Turbo codes) in order to protect the watermark. The casting of the watermark and other alternative modulation techniques are also analysed. In order to improve the system even further, 3-D marking replaces the usual frame by frame approach (2-D marking) by taking into account the temporal dimension. This increases the "local" chip rate leading to better cross-correlation results (wider cross-correlation area) and caters for frame dropping/duplication attacks.

Chapter 6 begins with an introduction to the wavelet transform, with the accent on the 2-D DWT (Discrete Wavelet Transform) case. The multiple advantages of the DWT transform are

discussed and compared with the traditional FFT/DCT transforms, taking into account the specific framework of digital watermarking. Choosing a proper basis constitutes an important step which will be also discussed. Due to major advantages of the DWT, the wavelet coefficients are one of the most suitable places to insert a watermark. The proposed watermarking system is described in detail during this chapter, including the HVS aspects of the scheme and error correction. The performance of the system will be then analysed for both image watermarking (in order to compare the results with the existing image watermarking schemes described in the literature) and video watermarking.

Chapter 7 discusses one of the most difficult problems in digital video watermarking: watermark recovery in the presence of geometric attacks like frame shift, cropping, scaling, rotation, and change of aspect ratio, especially when some of these are combined together. Re-establishing the synchronisation in a reasonable time is capital, and this is the object of this chapter. After a short introduction of the available techniques for combating geometrical attacks, this chapter proposes the use of an additional spatial watermark which combined together with image registration techniques will counteract the geometric attacks. The proposed technique, its implications and its performance are extensively analysed. The specific problems for video watermarking are also highlighted and measures to counteract these problems are proposed. As the results suggest this technique proves itself to be very successful, leading to a highly robust, high capacity blind watermarking system.

Chapter 8 will conclude the work presented in this thesis and suggest further research directions. The system's performance is summarised and compared with EBU's requirements. This comparison is presented in a tabular form, which fully illustrates the capabilities of the proposed system. In most of the cases, the proposed system meets or even exceeds (by far) the requirements of the EBU. Further research directions are also suggested during this chapter. This chapter also provides the complete list of author's publications.

Chapter 2

"I never found the perfect quote. At best I have been able to find a string of quotations which merely circle the ineffible idea I seek to express."

Caldwell O'Keefe

A Review of Watermarking

This chapter establishes the basis and specifies the requirements for the video watermarking system described in this thesis. The starting hypotheses of this project are detailed in close relationship with the EBU's video watermarking recommendations for a typical broadcast monitoring system. A brief overview of the existing watermarking techniques relevant to this thesis it is then presented, ending up by highlighting the contributions of this thesis.

2.1 A Basic Watermarking System

After a comprehensive survey of the existing techniques, by carefully analysing and weighting all the advantages and disadvantages of different methods a decision was reached to further pursue a spread spectrum technique, considering that spread spectrum has more potential compared with other methods, in spite of being so sensitive to the de-synchronization attacks.

Spread spectrum radio techniques have been developed for military applications, since mid 1940's for their anti-jamming and low-probability-of-intercept properties. They allow the reception of radio signals that are over 100 times weaker than the atmospheric noise.

In particular, the spread spectrum techniques are offering a good flexibility and are very suitable for watermarking due to the similarities between the watermarking and spread spectrum communications. Watermarking can be seen as a communication problem, in which

the original image plays the role of the channel noise and attackers may try to disrupt the transfer of information. In both cases the channel is a very difficult one characterised by high levels of noise. The large bandwidth required by a spread spectrum technique is not a problem, since usually the video sequences are quite big, offering a large number of coefficients and therefore the chip rate is sufficiently high for obtaining a robust watermarking system. The noise like spread spectrum signal is very difficult to detect/intercept and jam and is obviously spread in the entire video sequence, therefore suggesting a good robustness to certain attacks and a very secure system. Furthermore, the system can be relatively easy implemented, the watermark embedding and retrieving are based on secret keys and the system doesn't require the presence of the original video for watermark retrieving. A general block diagram for a video watermarking system based on the DSSS (Direct Sequence Spread Spectrum) is presented in **Figure 2-1**. The secret key is used for generating the same PN sequence for both embedding and retrieving. The spreading is achieved by multiplying this PN sequence with the data payload. As a result each watermark data bit is randomly spread in the entire video sequence, with a chip rate c_r. Typical for a video watermarking system, the recovery of the mark is blind, e.g. without resorting to the original video. The watermark is recovered by using

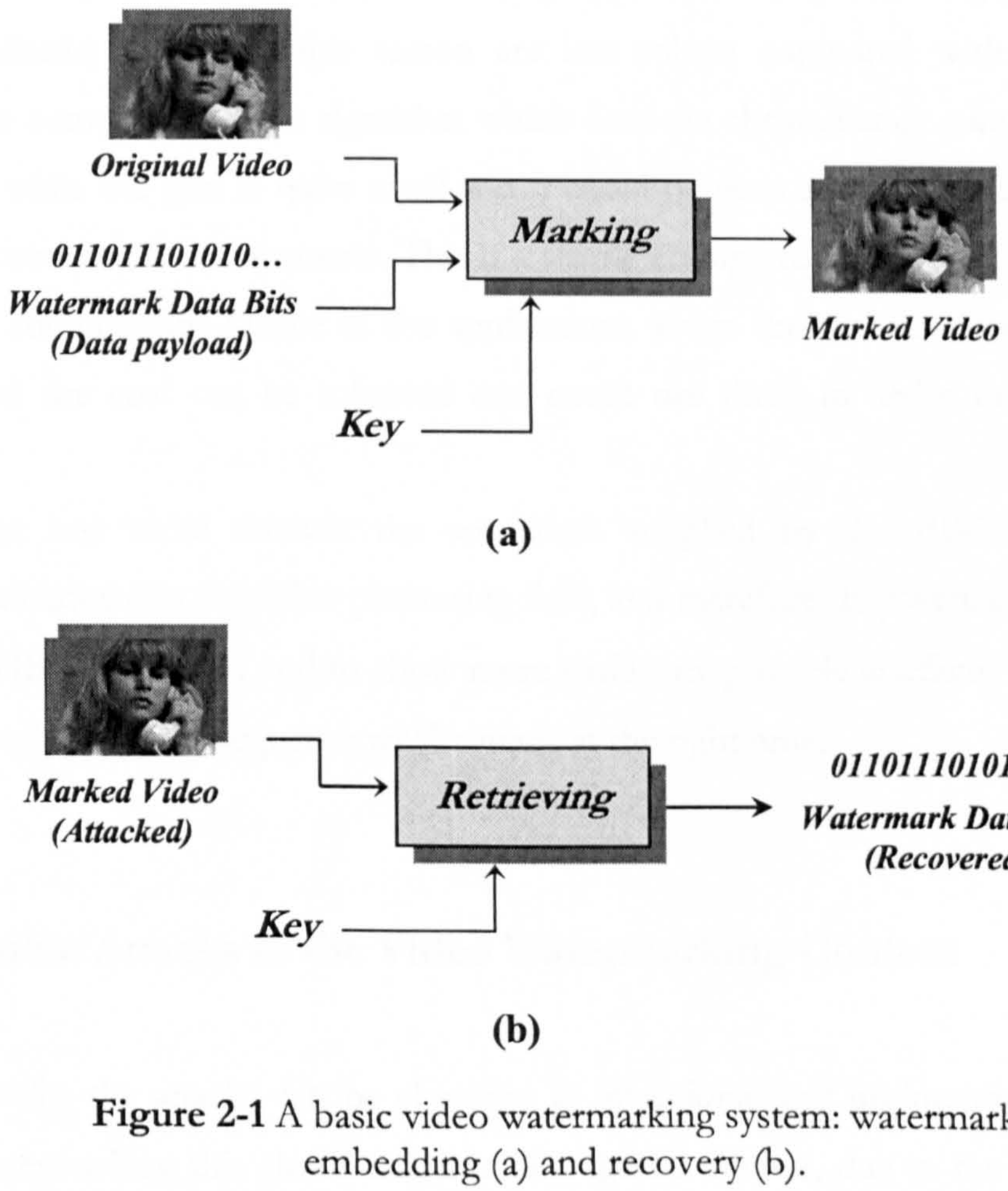


Figure 2-1 A basic video watermarking system: watermark embedding (a) and recovery (b).

cross-correlation methods, in the form of a matched filter (correlation receiver), following the principle of optimum reception. The SS technique will be described in detail in Chapter 3.

2.2 The Video Format

All the work done in this thesis was carried out in the context of uncompressed video, as found in TV studios and described by the ITU-R 601 (ITU-T BT.656) standard. To be more specific, the video sequences supplied by the BBC are in the raw Y-C_B-C_R format, with all the components separate and without any file header. The first half of the file represents the luminance component Y, followed by the chrominance components C_B and C_R.

The chrominance components are not robust at all, because they can be easily discarded, without affecting the video quality in any other way except the resulting black and white picture. Therefore all the algorithms described during the thesis are marking only the luminance component. Anyway marking the chrominance components has several other disadvantages. The human eye is much more sensitive to slight colour changes compared to slight luminance changes. As a result, these components have to be more lightly marked (with reduced amplitude) and from this reason are less robust compared with the luminance. Moreover, the complexity of the algorithm which uses the chrominance components is more than double, while the gain is quite small and it could be even zero if an attacker decides to discard the chrominance components. This is a strong enough reason to avoid the marking of chrominance components. Maybe in the applications where the real time requirement is not important and the cost can be tolerated one could use them in order to get a bit more robustness.

Finally, one more remark: the sequences supplied by the BBC are special test sequences, well known in the video processing field, and therefore they were especially selected to be more difficult to mark, and to show more easily any possible artefacts. The appropriate remarks for each particular sequence will be made at the right time.

2.3 Possible Attacks in the Video Watermarking Context

Generally, the attacks can be classified as intentional and un-intentional ("friendly"). For video watermarking this classification is of a special interest, due to the particularities of

the video processing chain. For example attacks like MPEG2 compression or slight spatial and/or temporal frame shifts can be considered un-intentional since they can appear during the video editing chain and their intent is not to destroy the watermark. In order to characterise an attack as intentional or unintentional, one has to carefully analyse the context of the application. Some attacks could easily fall in both categories, the difference being made by the intensity of the attack.

Although generally speaking each author has more or less his view when classifying the attacks, usually they can be divided in 3 main classes, as follows:

Signal processing attacks, which are probably the most usual category of attacks, contains mainly:

- signal enhancements: brightness, contrast, sharpening, blurring, etc
- digital and analogue, linear and non-linear filtering
- addition of noise and/or noise reduction
- digital-analogue (D/A) → analogue-digital (A/D) conversion and re-sampling
- data compression: MJPEG, MPEG1, MPEG2, DV, digital recording, etc
- PAL coding and analogue recording (VHS)
- colour space conversion / grey-scale conversion

Geometric attacks are one of the most efficient and demanding attacks against many watermarking systems. The most usual geometric attacks are:

- line/column cut and/or duplicate
- frame cut and/or duplicate
- frame rate conversion: 24 Hz ↔ 25Hz ↔ 30Hz
- picture aspect-ratio conversion: 4:3 ↔ 16:9
- line-scan conversion: progressive ↔ interlaced
- cropping, shifting (translation), rotation, scaling and possible others

Collusion and collusion-like attacks are attacks that use several copies of the same host media with different embedded watermarks, with the express purpose of removing the watermark. These copies are averaged together and a new data set (media) is created by using several different algorithms [Cox et al, 1995 and 1998], [Kilian et al, 1999], [Stone, 1996], [Craver et al, 1996, 1997 and 1998].

The ownership ambiguity attacks are trying to create an ambiguity about who's the real owner of the watermarked media. The so called "IBM attack" was first described in [Craver et al, 1996, 1997 and 1998] who suggests few techniques for combating this attack, like time-

stamps and non-invertible watermarks based on one way hashing functions (OWHF). At the moment these attacks are more or less forgotten, being considered as of second interest, especially in the video context; the EBU recommendations are not even mentioning this kind of attack.

These attacks and the robustness requirements for each class of attacks in the context of a broadcasting monitoring system are described in the EBU's watermarking recommendations [Cheveau et al, 2000]. It is worth mentioning here that these recommendations were published in March 2001, together with the test results of four watermarking systems provided by leading industry players: Lucent, Philips, Tektronix and Thomson, and this was the result of a "Watermarking Call for Systems" issued by EBU in May 2000. The conclusion of the tests was that none of the systems satisfied all the robustness requirements and only two of them complied with the 64 data bits capacity requirement. This shows once more that is not easy at all to comply with these requirements. Even at this time, none of the available systems can satisfy all these recommendations. To show once more the complexity of the problem, the conclusions drawn from the tests were that further development is needed together with a possible reduction of the data capacity from 64 to 48 data bits, in order to improve the robustness of the system [Cheveau et al, 2000].

2.4 Overview of the Existing Watermarking Techniques

In contrast with the steganography which has a very long history, the digital watermarking is a relatively young field, less than 10 years old. Practically, in the context of image/video watermarking the first papers appeared in the 1994, the debut being made in [Matsui et al, 1994] and [Schyndel at al, 1994]. They set the basis for the so called LSB (Least Significant Bit) watermarking. While this technique works in a noise free environment, it is totally useless when comes to robustness. Indeed, one can set the LSB either to 0 or to 1, without affecting the image quality too much and obviously the watermark is completely removed. In spite of this drawback quite a few variants were developed during 1994 and 1995.

The real breakthrough came in 1995/1996 with the introduction of spread spectrum watermarking [Cox et al, 1995 and 1996]. This can be regarded as the real starting point of robust watermarking. As expected, the first watermarking techniques were developed in the spatial domain, immediately followed by the DCT domain watermarking. Beside image

watermarking, which dominates the watermarking literature, we can find few uncompressed video watermarking techniques and several MPEG2 bit-stream watermarking techniques.

During the years several trends can be identified, starting with the spatial domain watermarking, closely followed by the DCT based watermarking. Developing of robustness benchmarking tools and watermarking algorithms which can withstand the attacks produced by these tools is another well represented trend in the watermarking research. The need of more robust systems led to introduction of perceptual watermarking, where the watermark is embedded according to HVS models. Few other research areas can be identified: the use of different transforms like FFT, DWT and several others, developing of RST (Rotation, Scaling and Translation) invariant watermarking techniques and finally the most recent trend is to regard the watermark as a hidden communication channel which is therefore protected by different error correction codes. Closely related with the communication theory and statistical modelling of the channel is the use of optimum detection theory in order to improve the reliability of the watermark detector. Hypothesis testing was extensively used in the detection process. Since the watermarking literature is quite large, only those papers directly relevant to this thesis are presented, describing briefly the algorithm for the most important ones.

2.4.1 Spatial Domain Watermarking Techniques

This is the first and the most straightforward way to add a watermark in an image/video. Ignoring the early schemes based on LSB, several pre-spread-spectrum image watermarking techniques can be mentioned: the "Patchwork" method in which randomly selected pairs of pixels are used to hide 1 bit by increasing the value of one pixel and decreasing the other one and the "Texture Block Coding" which embeds the watermark by copying one image texture block to another area in the image with a similar texture. Both methods were proposed in [Bender et al, 1995]. A similar "Patchwork" type technique, which this time divides the image into two equal sets was proposed in [Pitas et al, 1995 and 1996]. An improved version was later proposed in [Langelaar et al, 1996 and 1997]. Many other variations can be mentioned here but since they are not directly relevant to this thesis, the interested reader could find them in one of these watermarking overviews: [Kobayashi, 1997], [Swanson et al, 1998-1], [Hartung et al, 1999-2], [Wolfgang et al, 1997 and 1999] and [Langelaar et al, 2000].

Spread-spectrum and watermarking

As stated before one of the best way to insert a watermark is to use a technique based on spread-spectrum. The spread spectrum watermark is embedded in the media by "amplitude modulation". This is simply done by adding the watermark to the luminance values of the pixels. The entire mechanism is detailed in Chapter 3. Due to the simplicity and flexibility of the algorithm there are many watermarking schemes based on this principle.

One of the first and most well known representatives is the uncompressed video watermarking scheme developed by Hartung [Hartung et al, 1996 and 1998], who transposes the idea of Cox [Cox et al, 1995 and 1996] in spatial domain. The main advantage of Hartung's method is the blind recovery of the watermark. Another important difference is the multi-bit nature of the watermark, compared with Cox's scheme. The scheme uses a binary pseudo-sequence and uniform marking, e.g. in absolute value, each pixel is modified with the same value. Therefore the algorithm is non media dependent and non perceptual, which is indeed a significant drawback. In the latter paper, Hartung improves its algorithm by filtering the video sequence prior to cross-correlation and analyses the recovery of the watermark in more detail.

The "Watercast" system

Although it is widely accepted that the spatial domain is not the best place to cast a highly robust watermark [Ramkumar et al, 1998-2], [Fei et al, 2001], from the reasons discussed in the further chapters, one of the best video watermarking schemes available at this moment, Watercast [op de Beeck et al, 2001], [Kalker et al, 1999-1 and 1999-2] developed by Philips, uses the spatial domain.

Watercast is an improved version of JAWS (Just Another Watermarking System), adapted to the broadcast monitoring applications [Kalker et al, 1999-1 and 1999-2]. JAWS was initially developed for DVD copy-protection [Maes et al, 2000].

For the sake of maintaining low complexity, both watermark embedding and detection are performed in the spatial domain. The embedded watermark consists of watermark patterns of size 128x128 with Gaussian distribution, which are repeated (tiled) to fill the whole video frame. In order to avoid visible artefacts, the watermark is scaled on a pixel-by-pixel basis, with a scaling factor which is derived from an "activity measure". The "activity measure" is in fact an empirical HVS model, computed using a Laplacian high-pass filter. This is quite a common method in digital watermarking. More details about this type of HVS marking can be found in Chapter 3. The same watermark is embedded into several consecutive video frames.

For watermark detection, a correlation detector is used after applying a spatial pre-filter [Depovere et al, 1998] that reduces cross-talk between video signal and watermark. Since the watermark must be detected even in the presence of spatial shifts, a search over all possible shifts is performed. Because the watermark signal is generated by tiling of a smaller watermark pattern, only 128 x 128 positions have to be searched, according to the size of the watermark pattern. In order to reduce complexity, the search and correlation is done in the FFT domain. Further, only the phase information of the FFT is used in the correlation. This method of detection has been previously proposed for pattern recognition and is referred to as Symmetrical Phase Only Matched Filtering (SPOMF) [Kuglin et al, 1975], [Chen et al, 1994] and [Pech-Pacheco et al, 199x].

In order to embed a sufficiently large multi-bit watermark, the system has to use several different basic watermark patterns on top of each other. The information is encoded in the choice of the basic patterns and their relative positions, leading overall to a quite complex system. The watermark can convey just enough information to comply with the EBU's requirements.

Other systems

Coming back to the image watermarking, several other authors could be mentioned: [Swanson et al, 1996-1], [Nikolaidis et al, 1998], [Kutter, 1999-2] and [Queluz et al, 2000]. Few other techniques can be found in [Kobayashi, 1997], [Swanson et al, 1998-1], [Hartung et al, 1999-2], [Wolfgang et al, 1999] and [Langelaar et al, 2000].

Blind versus non-blind recovery

Browsing the literature, it can be remarked that initially most of the watermarking techniques were non-blind, requiring the use of the original in the detection process. Later these techniques were adapted to work with blind recovery. Most of the watermarking techniques used the entire image/frame without dividing it into small blocks, later evolving into block based schemes.

Perceptual considerations

From the perspective of watermark embedding, most of the early techniques relied on uniform marking, with an amplification factor experimentally tweaked for acceptable visibility.

The HVS based marking started to appear in the form of region classification, different edge/gradient detection algorithms and different other empirical measures of the local activity within a block/area. It must be remarked here that all these methods are more or less empirical, due to the lack of HVS models in the spatial domain.

Capacity of the watermarking system

From the capacity point of view, most of the watermarking schemes were capable of hiding only one data bit, or in other words they were only capable to tell if a certain watermark was found or not in the media. Quite late, the multi-bit schemes finally arrived. Another major improvement which surprisingly was adopted quite late was the use of pre-filtering as a mean of reducing the cross-talk between the original media and the watermark.

2.4.2 Watermarking in the DCT Domain

The DCT domain is far the most popular one, from several reasons. One reason is that all the major compression techniques were developed in the DCT domain (JPEG, MJPEG, MPEG1, MPEG2, H26x) and therefore the image processing community was familiar with it. Much research was carried out in developing various perceptual models for the DCT domain, and these models could be easily applied to watermarking, since watermarking and compression are very closely related. Since the compression algorithms are well known, one could compensate for it during the watermark embedding process, making the algorithm robust against compression. Furthermore marking in the frequency domain rather than spatial domain has few advantages: better robustness against certain attacks, higher capacity, more close to the HVS and relatively good frequency localisation of the coefficients. Those who are marking in the bit-stream domain (MPEG2) have the additional advantage of the direct bit-stream marking, without decoding and re-encoding the signal.

Spread-spectrum watermarking

One cannot start talking about DCT domain watermarking without mentioning from the beginning the image watermarking scheme developed by Cox [Cox et al, 1995 and 1996]. As mentioned before, Cox officially introduced the use of spread spectrum in digital watermarking, being one of the most cited authors in the watermarking world. In their scheme, the watermark is inserted in the DCT domain. Cox was probably between the first to realise the importance of perceptual marking, and although they do not actually use a visual model,

the watermark is embedded in what they regarded to be the most relevant regions of the signal. Therefore, the DCT transform of the entire NxN image was computed and the watermark was inserted in the first n highest magnitude DCT coefficients, excluding the DC coefficient. Due to this arrangement, the necessity of the original image for watermark detection is obvious, which is a serious drawback. The scheme basically embeds only one bit, which is recovered by computing the similarity (which is in fact another name for normalised cross-correlation) between the original and extracted watermarks.

Full-frame versus block-based embedding

Similar with the spatial domain schemes, one can identify techniques which are embedding the mark in the full-frame DCT coefficients [Cox et al, 1995 and 1996], [Barni et al, 1998], [Bartolini et al, 1998-1], [Piva et al, 1998] or in the block-wise manner [Swanson et al, 1996-1 and 1996-2], [Podilchuk et al, 1997-1, 1997-2 and 1998], [Wolfgang et al, 1999], [Ramkumar et al, 1998-1], [Kim et al, 1999], [Zhu et al all, 1996], [Tao et al, 1997] and [Hernandez et al, 1998-2, 1999-1 and 1999-2]. The main reason for adopting the block-based approach is robustness to certain geometrical attacks. It is also helpful that most of the existing HVS models are working on a block-based basis.

HVS-based marking

The watermarking research community realised pretty soon that uniform marking is not the best choice of embedding a watermark, since offers quite an unfavourable robustness/visibility report. Inserting a watermark tacking into account some aspects of human vision led to better results, as more energy can be packed into the media while keeping the visibility of the watermark low. After different trials with some more or less empirical methods, finally more advanced HVS models mostly developed in the context of human/machine vision and image/video compression started to be adapted to the needs of watermarking.

Papers like [Carlson et al, 1980], [Legge et al, 1980], [Girod, 1989], [Peterson et al, 1993], [Ahumada et al, 1992], [Jayant et al, 1993-1 and 1993-2], [Watson, 1993], [Watson et al, 1994], [Zhu et al, 1995], [Chou et al, 1995 and 1996], [Eckert et al, 1998] started to become actual and the models described started to be adapted and applied to watermarking, leading to important robustness gains.

Cox is underlining the importance of perceptual embedding in [Cox et al, 1997]. Some of the most important examples of watermarking systems incorporating HVS models are mentioned below.

The watermarking system described in [Swanson et al, 1996-1, 1996-2 and 1997] and [Zhu et al, 1996] is based on the HVS model described in [Zhu et al, 1995]. This is in fact a modified version of the TEL (Tolerable Error Level) model developed by Girod [Girod, 1989]. This rather complex model is based on both spatial and frequency domain masking models.

Some authors preferred to develop their own algorithms, for example [Tao et al, 1997], [Delaigle et al, 1998] and [Bartolini et al, 1998]. Tao used a regional perceptual classifier, which assigns noise-sensitivity indexes to each DCT block. The algorithm exploits luminance, edge and texture masking effects of the HVS and classifies a block into one of 6 categories [Tao et al, 1997]. Following a similar approach, [Bartolini et al, 1998] uses a combination of several different filters and thresholds to build a perceptual mask.

In [Podilchuk et al, 1997-1, 1997-2 and 1998] and [Wolfgang et al, 1999] the authors use a modified version of the JND (Just Noticeable Distortion) model developed at NASA [Peterson et al, 1993], [Ahumada et al, 1992], [Watson, 1993] and [Watson et al, 1994]. It is largely believed that the JND model (sometimes called Watson's model) is the best DCT based HVS model available. This is due to the highly adaptive nature of this model which takes into account the most important masking effects of the HVS: the MTF (Modulation Transfer Function) of the eye, the luminance and the contrast masking. The model is capable to accurately estimate a JND level for each DCT coefficient within a block in contrast with the other HVS models which usually assign only one threshold for the entire block. [Kim et al, 1999] improves the scheme described in [Podilchuk et al, 1997-1, 1997-2 and 1998] and [Wolfgang et al, 1999] by extending the JND model to account for another HVS masking effect: lateral inhibition masking.

Several other papers describing watermarking systems based on HVS models could be quoted and few other visual models as well, but they are not directly relevant with the work described in this thesis and therefore they will not be presented here. The interested reader could find them in these comprehensive watermarking reviews: [Kobayashi, 1997], [Swanson et al, 1998-1], [Hartung et al, 1999-2], [Wolfgang et al, 1997 and 1999] and [Langelaar et al, 2000]. A review of existing visual models could be found in [Jayant et al, 1993-1, 1993-2] and [Eckert et al, 1998].

Detection (recovery) of the watermark

Many techniques are requiring the original in the detection process [Cox et al, 1995 and 1996], [Swanson et al, 1996-1], [Tao et al, 1997], [Podilchuk et al, 1997-1 and 1997-2], [Wolfgang et al, 1999] and [Kim et al, 1999] while many other are blind [Swanson et al, 1996-2],

[Barni et al, 1998-3], [Bartolini et al, 1998], [Piva et al, 1998], [Ramkumar et al, 1998-1], [Wolfgang et al, 1999] and [Zhu et al, 1996].

For example in [Barni et al, 1998-3] the authors use a modified form of Cox's technique which does not require the original for recovery. They insert the watermark in a known fixed location: the coefficients from the $(L+1)th$ to the $(L+M)th$ are taken according with the zigzag ordering of the DCT spectrum, where the first L coefficients are skipped to achieve perceptual invisibility of the mark.

Speaking by watermark recovery, much work was carried during the time in order to improve the reliability of the detectors and to develop better detection models and strategies. In particular, many researchers used the existing optimum detection theory developed in the context of communications [Hernandez et al, 1998-2, 1999-1, 1999-2 and 2000-1], [Barni et al, 1998-1 and 1998-2], [Piva et al, 1998 and 2000], [Linnartz et al, 1997], [Robert et al, 2000].

Watermark embedding techniques

Other researchers concentrated their efforts to improve the "modulation" techniques, or in other words searched for better ways of casting the watermark.

For example [Smith et al, 1996] suggested the use of differential modulation techniques in watermarking. On the same note, [Lu et al, 1999 and 2000] proposed a scheme called "Cocktail watermarking" which embeds one data bit in the signs of two coefficients or blocks. They defined four possible types of modulations: $Modu(+,+)$, $Modu(+,-)$, $Modu(-,+)$ and $Modu(-,-)$, where $Modu(+/-,-/+)$ represents a positive/negative transformed coefficient modulated with a negative/positive watermark quantity.

Analysing the influence of a number of attacks in order to see how the coefficients are modified, the authors claim that different attacks are affecting the magnitude of the coefficients in a biased way, and therefore by using the appropriate form of modulation (positive modulation or negative modulation function of the attack) the detector response increases. The embedding is based on a HVS model but the capacity of the scheme is only one data bit. Their technique was applied to both DCT and Wavelet coefficients.

Communication theory and capacity

Most of the early techniques are capable of inserting only one data bit [Cox et al, 1995 and 1996], [Swanson et al, 1996-1], [Tao et al, 1997], [Podilchuk et al, 1997-1 and 1997-2], [Wolfgang et al, 1999], [Barni et al, 1998-1, 1998-2 and 1998-3], [Bartolini et al, 1998] and [Kim

et al, 1999] while most of the new techniques are capable of embedding a multi-bit watermark [Swanson et al, 1996-2], [Bami et al, capacity 1999-1], [Perez-Gonzales et al, 2001], [Ramkumar et al, 1998-2 and 1999] and [Hernandez et al, 2000-2].

In 1996, Smith and Comisky raised the capital question: How many bits can we hide into an image? They showed for the first time [Smith et al, 1996], using communication theory and in particular Shannon's channel capacity that the maximum capacity of an image is quite large. This is remarkable bearing in mind that the watermarking was only at its beginnings.

Error correction codes and watermarking

Several years had to pass, until the watermarking research community pushed by the need of more robust watermarks started to see the watermarking as a hidden communication channel which therefore can be protected by FEC (Forward Error-correction Codes) in order to improve the robustness [Mittelholzer, 1999], [Ramkumar et al, 1998-2 and 1999], [Cox et al, 1999], [Perez-Gonzales et al, 2001], [Hernandez et al, 1998-1, and 2000-2], [Barni et al, 1999-1], [Baudry et al, 2001], [Moulin et al, 2001] and [Fei et al, 2001]. The choice of the codes ranges from the most basic ones to the powerful Turbo codes. The use of coding led to significantly better results. Once with the introduction of FEC, the performance of the watermarking schemes begun to be measured in terms of BER (Bit Error Rate).

Video watermarking

Excepting the MPEG2 bit stream watermarking, which does not make the object of this investigation, and in spite of few papers having the word "video" in their title but which in fact are not dealing with the video at all, the DCT-based video watermarking schemes are almost inexistent.

The exception from the rule is the object-based scheme described in [Swanson et al, 1997] which uses segmentation algorithms and HVS models, in a block-based approach. The technique used in their scheme is similar with the MPEG motion tracking. The HVS model is the rather complex TEL model developed by Girod [Girod, 1989]. The results of the scheme are not very bad, but the scheme has a major drawback: it is capable to embed only one single data bit.

Another exception is [Busch et al, 1999] where the authors applied an already known DCT block-based technique developed for still-image watermarking to video. The scheme embeds the watermark only in the luminance component. In order to improve the invisibility of the watermark, the embedding is performed only in those blocks qualified as appropriate by

a block activity measure. This is quite a poor choice of HVS model, and as a result the performance of the scheme is rather weak. The authors suggest the use of frame averaging in order to increase the robustness of their scheme, but even so, overall the scheme gives only poor results. Even when 64 data bits are embedded in 50 consecutive frames, the scheme still yields quite a high BER, and the authors suggest using an even longer watermark segment.

2.4.3 Fourier Domain Watermarking

Obviously the DFT (Discrete Fourier Transform) was used for watermarking as well. Unlike in the spatial or the DCT domains, the number of papers dealing with DFT marking is quite low. Each transform domain has its own advantages and disadvantages and the DFT is a particularly good example.

The DFT is shift (translation) invariant, or in other words cyclic shifts of the image in the spatial domain do not affect the magnitude of the DFT and therefore a watermark embedded in the DFT domain will be shift invariant. Of course this is a highly desirable property.

On the other hand, as [O'Ruanaidh et al, 1996] shows, due to its complex nature, the DFT offers the possibility of watermarking the magnitude (amplitude) or the phase (at least in theory). The phase is far more important than the magnitude of the DFT values for the intelligibility of an image, so embedding a watermark in the most important component of an image is very good since any attempts of removing the watermark will lead to heavy artefacts. Moreover, as known from the communication theory, the phase modulation often possesses superior noise immunity in comparison with amplitude modulation.

Based on these observations, in [O'Ruanaidh et al, 1996] the authors decide to mark the phase rather than the magnitude of the DFT coefficients. Well, this idea was quickly dropped, and never pursued again, without explaining the reason. Experiments show that probably one of the reasons was the sensitivity of the phase to JPEG and MPEG attacks.

On the other hand, another major disadvantage of both phase and magnitude marking is the fact that in order to obtain a real image after IDFT, the following symmetry conditions must be fulfilled: changes in magnitude must preserve the positive symmetry of the Fourier coefficients and changes in phase must preserve the negative symmetry of the Fourier coefficients. These symmetry requirements are basically halving the watermarking space and therefore the capacity, being a serious drawback. Furthermore, the lack of HVS models in the Fourier domain is another drawback of the FFT-based watermarking.

If the phase marking is almost inexistent in the literature, the magnitude marking instead has a better faith [Licks et al, 1999 and 2000], [Solachidis et al, 1999], [Piva et al, 2000], [Ramkumar et al, 1998-1 and 1999], [Deguillaume et al, 1999].

In an attempt to exploit the DFT properties, papers like [Solachidis et al, 1999] propose to embed the watermark in a circularly symmetric manner. The obvious advantage is the robustness to certain geometrical attacks like cropping, shifting and rotation for example, but the capacity of the scheme is quite low.

In a rare approach, described in [Deguillaume et al, 1999], the authors propose to embed a spread spectrum watermark into 3-D DFT blocks of video, by employing a 3-D DFT and adding the watermark to the transform coefficients. Additionally they embed a template which is easy to detect even under geometric attacks, but overall the scheme gives only modest results.

Although a shift in the spatial domain does not affect the magnitude of the Fourier coefficients, it will instead affect the phase: a shift in spatial domain is equivalent with a phase shift in the Fourier domain. Furthermore, according to the convolution theorem, cross-correlation in spatial domain is equivalent with multiplication in the FFT domain, and vice versa. As a result, the FFT transform is often used for implementing fast cross-correlators. Since a sliding correlator (e.g. a cross-correlator which is able to search for the right position of the watermark in an attacked image) is very computationally expensive (section 3.2.2), the efficiency of the FFT correlators is particularly welcomed. An example of such a correlator is SPOMF (Symmetrical Phase Only Matched Filter), which was already mentioned and which will be used in Chapter 7.

Another particular case which involves the use of Fourier transform is represented by the RST (Rotation, Scaling and Translation) invariant watermarking schemes [Kutter, 1998], [O'Ruanaidh et al, 1998], [Deguillaume et al, 1999] and [Lin et al, 2000]. This case will be discussed in Chapter 7.

2.4.4 Watermarking in the Wavelet Domain

As the watermarking literature suggests and this thesis confirms, the watermarking in wavelet domain is one of the best choices available. The wavelet transform is starting to become more and more popular, as the researchers are starting to be aware of the multiple advantages offered by this transform. These advantages are extensively presented in Chapter 6. Although some authors suggested the use of wavelet watermarking techniques few years before

this trend started to appear in the research community, as most of the ideas proposed ahead of their on time, they passed almost unnoticed.

Like any other watermarking scheme, a DWT-based scheme can be characterised by blind/non-blind recovery, one bit/multi-bit watermark and HVS/non-HVS watermark embedding. In the DWT context, the HVS-based embedding is not as critical as in the other domains, due to the fact that the wavelet transform is almost a HVS model by itself. For example the human eye is less sensitive to noise in high resolution DWT bands (level 1) and especially in the DWT bands having an orientation of 45° (i.e., HH bands).

Another motive for watermarking in wavelet domain is the JPEG 2000 standard, based on the so called embedded zero-tree wavelet coding (EZW), and which will replace sooner or later the old JPEG standard due to its reduced visibility artefacts and better compression. One major advantage compared with its closest competitor - the DCT compression/watermarking methods - is the absence of those very annoying blocking artefacts characteristic to all DCT compression/watermarking schemes. The reason for this big advantage is that DWT is not a block based transform. Furthermore, the DWT is easier to compute than the DCT. Another advantage related with the JPEG 2000 is the ability of tweaking the watermarking algorithm in such a way that becomes robust to JPEG 2000 compression.

The first paper to propose the use of wavelet transform was [Boland et al, 1995] followed by [Podilchuk et al, 1997-1 and 1998], [Swanson et al, 1998-2], [Inoue et al, 1998], [Wolfgang et al, 1999], [Kundur et al, 1997 and 1998], [Lin et al, 1998], [Jayawardena et al, 2000], [Lumini et al, 2000], [Loo et al, 2000], [Dugad et al, 1998], [Pereira et al, 2000], [Barni et al, 1999-2], [Lee et al, 2000], [Xia et al, 1998], [Wang et al, 1998] and [Tsekeridou et al, 2000].

HVS-based marking

It was mentioned before that using HVS models in wavelet domain is not as critical as in the other domains due to the similarity of the DWT with the HVS. Indeed research into human perception indicates that the retina of the eye splits an image into several frequency channels each spanning a bandwidth of approximately one octave. The signals in these channels are processed independently. With its multi-resolution nature, the wavelet transform separates the image into bands of approximately equal bandwidth on a logarithmic scale [Kundur et al, 1997].

In fact some authors suggested that a HVS model in not even necessary for DWT-based watermarking. Evidently this is not true, since a visual model can pack more energy and reduce the visibility of the watermark which translates to increased capacity and robustness of

the watermarking system, but some authors may be right to say that usually in the wavelet domain a simpler model may suffice.

Due to these considerations and because the wavelet transform is much younger than its counterparts and therefore the number of compression methods existent in the wavelet domain is quite low compared with those existing in the DCT domain, the choice of HVS models is much reduced.

The most important wavelet HVS models are those developed in [Lewis et al, 1992] and [Watson et al, 1996 and 1997]. Unlike the advanced JND model designed for the DCT by the same author [Watson, 1993], [Watson et al, 1994], the wavelet model is quite simple, giving only one quantisation factor for each wavelet sub-band. In contrast the Lewis model is much more adaptive and therefore much more complex. [Podilchuk et al, 1997-1, 1997-2 and 1998] was the first paper to introduce a proper HVS model. While the authors employ the simpler Watson model, other papers [Barni et al, 1999-2] are using the more complex Lewis model.

Watermark recovery and system's capacity

From the capacity perspective, amazingly enough, the vast majority of existing techniques are only capable of embedding one single bit. The only schemes capable of embedding a multi-bit watermark are those described in [Loo et al, 2000] and [Pereira et al, 2000]. In fact few schemes are so primitive that they are not even worth mentioning.

Most of the techniques are blind [Inoue et al, 1998], [Lumini et al, 2000], [Loo et al, 2000], [Dugad et al, 1998], [Pereira et al, 2000], [Barni et al, 1999-2], [Wang et al, 1998] and [Tsekeridou et al, 2000] but quite a few of them are requiring the original for detection [Podilchuk et al, 1997-1, 1997-2 and 1998], [Swanson et al, 1998-2], [Xia et al, 1998] and [Lee et al, 2000].

It is quite unbelievable that most of the existing wavelet techniques are either non-blind or are capable of embedding only one data bit. This aspect is even more puzzling considering that most of these techniques appeared quite recently. But the worse is still to come.

Video watermarking schemes

Uncompressed video watermarking, is a rarity in wavelet domain. Basically excluding those papers having "video" in their title just for creating an impression, the video watermarking in wavelet domain resumes to [Swanson et al, 1998-2] and [Lee et al, 2000].

The technique described in [Swanson et al, 1998-2] is quite unique. The authors employ a temporal wavelet transform along the video frames, and watermark the wavelet coefficients

in a similar manner with their DCT-based scheme presented in section 2.4.2. Unfortunately the wavelet scheme is even more complex than their DCT scheme and doesn't really take advantage of all the opportunities offered by the wavelet transform. The scheme uses both wavelet transform and the DCT transform of the wavelet coefficients, segmentation algorithms and the complex Girod HVS model, therefore being highly inefficient in terms of computing cost. But the real problem is that the scheme is both non-blind and capable of embedding only one data bit into the video sequence, rendering it more or less useless.

In a more recent scheme [Lee et al, 2000], the watermark is adaptively embedded using a HVS model. Moreover, in order to take advantage of the temporal dimension the algorithm take into account region complexity and motion information. Unfortunately this scheme is capable of embedding only one data bit and even worse, requires the presence of the original in the watermark detection process. The detection method used is the one proposed by Cox in his original image watermarking scheme. Therefore the scheme is highly inappropriate to any practical use and is pretty weak in terms of robustness as even the authors acknowledge.

As the case of image watermarking in the wavelet domain wasn't bad enough, even the video watermarking schemes are non-blind and capable of casting only one data bit, transforming them into a publishing exercise rather than a scheme which could be used in practice.

2.5 Contributions

The work presented in this thesis was partly fuelled by a comprehensive literature survey, performed at various stages during the PhD. An integral part of this work was the investigation of different existing watermarking techniques for spatial, DCT and DWT domain watermarking. One of the main conclusions of the literature survey was the importance of regarding and analysing the watermarking system from the communication perspective. Following this path, the watermarking channel is seen as a communication channel and the performance of the watermarking scheme is significantly improved by using powerful error correction codes (Turbo codes). Another main point drawn from this investigation was the importance of the HVS models in any watermarking system. Therefore, various existing visual models developed in the context of image compression were investigated. Simplifying, improving and adapting these models to the requirements of blind digital video watermarking - for both DCT and DWT systems - was an important step in achieving the final result.

This work begun with the developing of two spatial domain, blind video watermarking schemes used as an initial environment appropriate for investigating the effects of pre-filtering, block sizes and HVS models on the performance of a spatial domain watermarking scheme. By using this knowledge and the conclusions drawn from these initial spatial domain watermarking schemes, this work established the foundations for the following high performance, transform domain watermarking systems.

The main contributions of the thesis can be summarised as follows:

- Developing a blind, robust DCT-based video watermarking scheme based on an advanced HVS model, and incorporating state of the art FEC, 3-D marking and 3-D sliding correlation.
- Investigating the use of wavelet transform in digital watermarking and developing a high capacity, robust wavelet-based blind video watermarking system which takes advantage of the properties of the wavelet transform. The performance of the system is further improved by using HVS model and advanced FEC.
- Investigating different techniques against geometrical attacks and developing a system based on image registration techniques which combined with the existing wavelet scheme leads to a highly robust, high capacity, blind video watermarking system capable to withstand a wide range of geometrical attacks. This illustrates a new class of application for image/video watermarking: "blind video registration".

Chapter 3

"And if, to be sure, sometimes you need to conceal a fact with words, do it in such a way that it does not become known ..."

Nicollo Machiavelli (1469-1527)

Spatial Domain Watermarking

The chapter presents in detail the structure and the algorithm of a spread spectrum video watermarking system. This chapter establishes the foundation and the skeleton of the spread spectrum watermarking system. The main structure and many of the components described here are used for building the DCT scheme from Chapter 5. Starting with the basic uniform marking, the system is gradually improved by embedding the watermark in a more efficient way (e.g. HVS dependent) and by using a sliding correlator for watermark recovery. The effects of pre-filtering and various block sizes on the performance of a spatial domain watermarking scheme are also investigated.

3.1 Watermark Embedding

3.1.1 Embedding the Watermark: the Spread-Spectrum Approach

As already stated, the basic idea of spread spectrum is to trade signal-to-noise ratio (SNR) for bandwidth. In other words, the signal energy is spread over a wide frequency range at low SNR so that it is difficult to detect, intercept or jam. Though the total signal power may be large, the SNR in any band is small. Moreover the signal energy resides in all frequency bands.

For watermarking this translates to a visually imperceptible watermark, spread in the entire video sequence. Since the watermark resides in all frequency bands, it is likely that even when attacked at least some parts of the spectrum still remain intact, given that usually the attacks are band limited (for example compression, which removes the high frequency components of the spectrum). The spreading signal, which is in fact a pseudo-random or pseudo-noise sequence, who will be called "PN sequence" from now on, it is unknown to a potential attacker, being generated function of a secret key.

The DSSS (Direct Sequence Spread Spectrum) is the simplest form of a spread spectrum technique. Basically the signal is modulated by a function that alternates pseudo-randomly between $+\alpha$ and $-\alpha$, at multiples of a time constant called the *chip rate*. This pseudo-random carrier contains components of all frequencies, which is why it spreads the modulated signal's energy over a large frequency band. For watermarking, the chip rate can be considered as the spacing between the pixels.

Let's assume that $u_k, u_k \in \{-1,1\}$ are the *input data bits* or the *payload* which is to be hidden into the video sequence, converted from the $\{0,1\}$ values to $\{-1,1\}$ values. This signal is spread by a large chip rate factor c_r to give the spread sequence b_i

$$b_i = u_k, \quad k \cdot c_r \le i \le (k+1) \cdot c_r \tag{3.1}$$

Each data bit now spans c_r binary bits. The spread sequence b_i is amplified with an amplitude adjustment factor α and then modulated with the binary PN sequence p_i, $p_i \in \{-1,1\}$ giving the watermark signal

$$w_i = \alpha \cdot p_i \cdot b_i \tag{3.2}$$

Finally the watermark is added to the luminance component of the video sequence v_i, giving the watermarked signal

$$v_i^M = v_i + w_i = v_i + \alpha \cdot p_i \cdot b_i \tag{3.3}$$

Some authors are calling this "amplitude modulation".

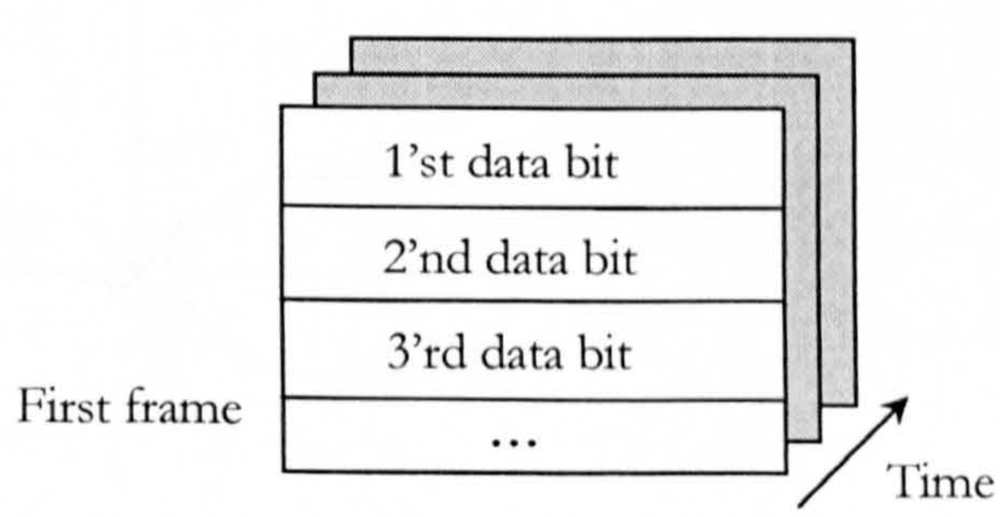


Figure 3-1 Hartung's method of spreading the payload

This algorithm is used in [Hartung et al, 1996 and 1998]. From the embedding side, the algorithm has two major weaknesses. Firstly the algorithm uses uniform marking (a constant amplification factor for the entire video sequence) and therefore is not adaptive in the sense of incorporating at least some HVS aspects. This leads to reduced robustness especially in the case of compression. Secondly, the way described in equation (3.1) for spreading the payload is not the most secure and robust way to do it, since all the data bits are grouped together in a region of the video and embedded one group after another. A possible case is illustrated in

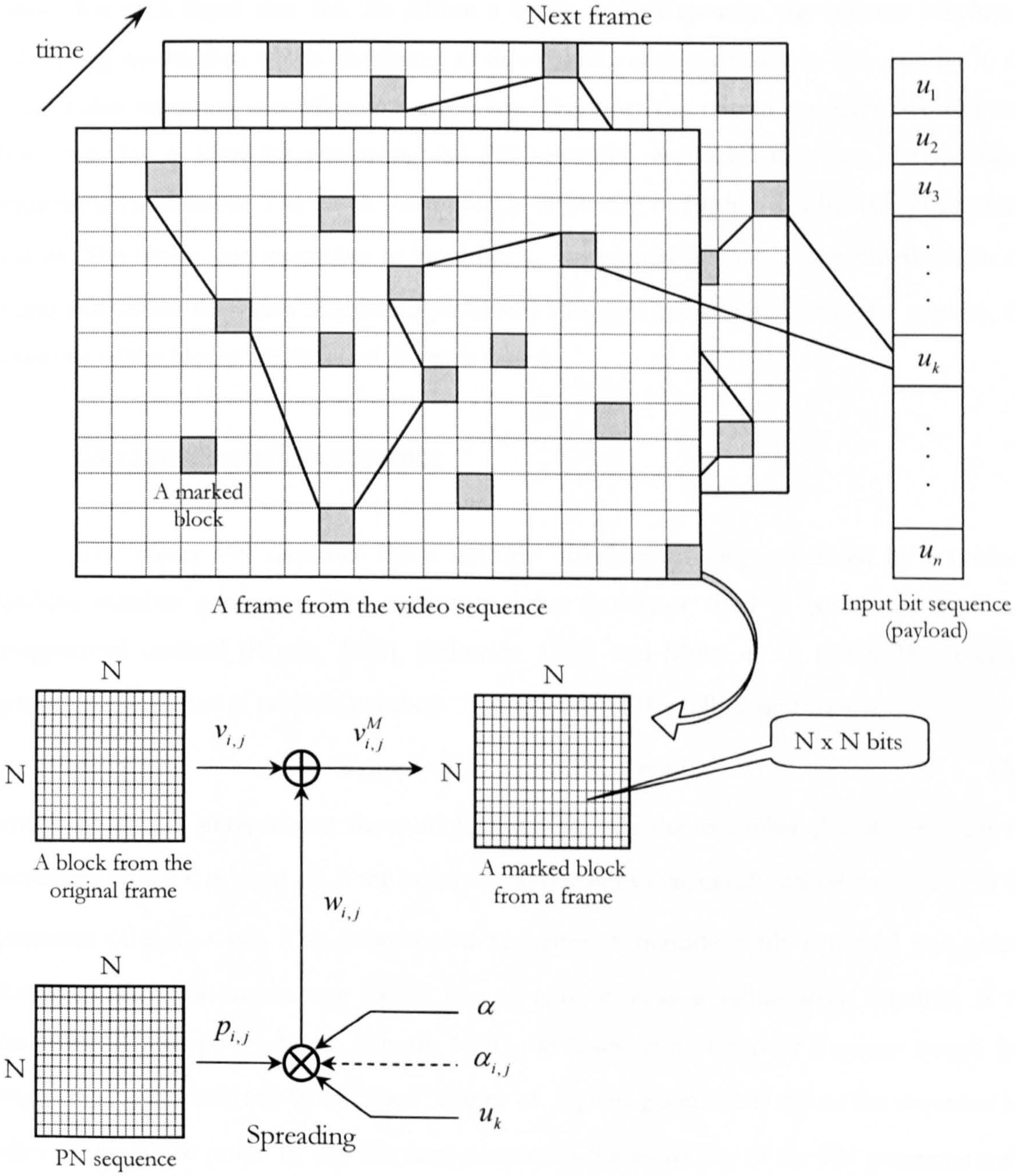


Figure 3-2 Secure, block-based video watermarking – the embedding process

Figure 3-1. In this case if an attacker removes the first frame, it is impossible to recover the first 3 bits (and possibly the next one) even by using a sliding correlator, since because of their spatial localisation they are lost. A much better way is to spread the sequence in the entire video.

The system presented in **Figure 3-2** was designed having these considerations in mind. From reasons of security and robustness to geometric attacks like line cuts or column cuts, the scheme is block based rather than full frame. Each NxN block corresponds to one data bit from the input sequence. Overall, due to the spreading the algorithm assigns a number of NxN blocks for each input data bit. To ensure a better system security, the scheme employs a scrambling mechanism which distributes all these blocks corresponding to one data bit in the entire video sequence, according to a given key. Therefore the system can have two different keys: one key is used for generating the PN sequence, and the other one for the block scrambling mechanism. The block scrambling is illustrated in **Figure 3-2** for two consecutive frames. The blocks corresponding to input bit u_k are pseudo-randomly distributed within the frame and within the video sequence (spatial and temporal spreading). For better security, the locations of the blocks are different for each frame.

The pseudo-random sequence

The binary PN sequence has a uniform distribution being generated by a uniform random number generator. The random number generator used is based on the linear congruential method [Knuth, 1981], [Schneier, 1996] and [Press et al, 1992]. This method generates a sequence of random numbers X_n according to the following formula

$$X_{n+1} = (aX_n + c) \bmod m, \quad n \geq 0 \tag{3.4}$$

where the integer m represents the modulus ($m > 0$), a is the multiplier ($0 \leq a < m$), c is the increment ($0 \leq c < m$) and X_0 is the initial or the starting value, usually called the "seed" of the generator ($0 \leq X_0 < m$). The congruential sequence is periodic with a period not greater than m. From this reason one would like to choose m as a rather large number. If the multiplier a is properly chosen [Knuth, 1981] this leads to a period of maximal length (e.g. length m). In this case any initial "seed" choice of X_0 is as good as any other: the sequence just takes off from that point. In fact this seed constitutes the secret key of the PN generator and is a 32 bit integer.

The special case $c = 0$ leads to a faster algorithm but reduces the length of the period of the sequence. Nevertheless it is still possible to make the period sufficiently long. This case is known as the multiplicative congruential method or mixed congruential method. The generator chosen for this work is an improved multiplicative congruential pseudo-random generator [Press et al, 1992]. The algorithm uses the L'Ecuyer method with Bays-Durham shuffle and added safeguards and has a period greater than 2×10^{18}. [Press et al, 1992] calls this the "perfect" random number generator offering to pay 1000USD to the first reader who can prove the contrary. In terms of security, this generator is considered as being relatively insecure.

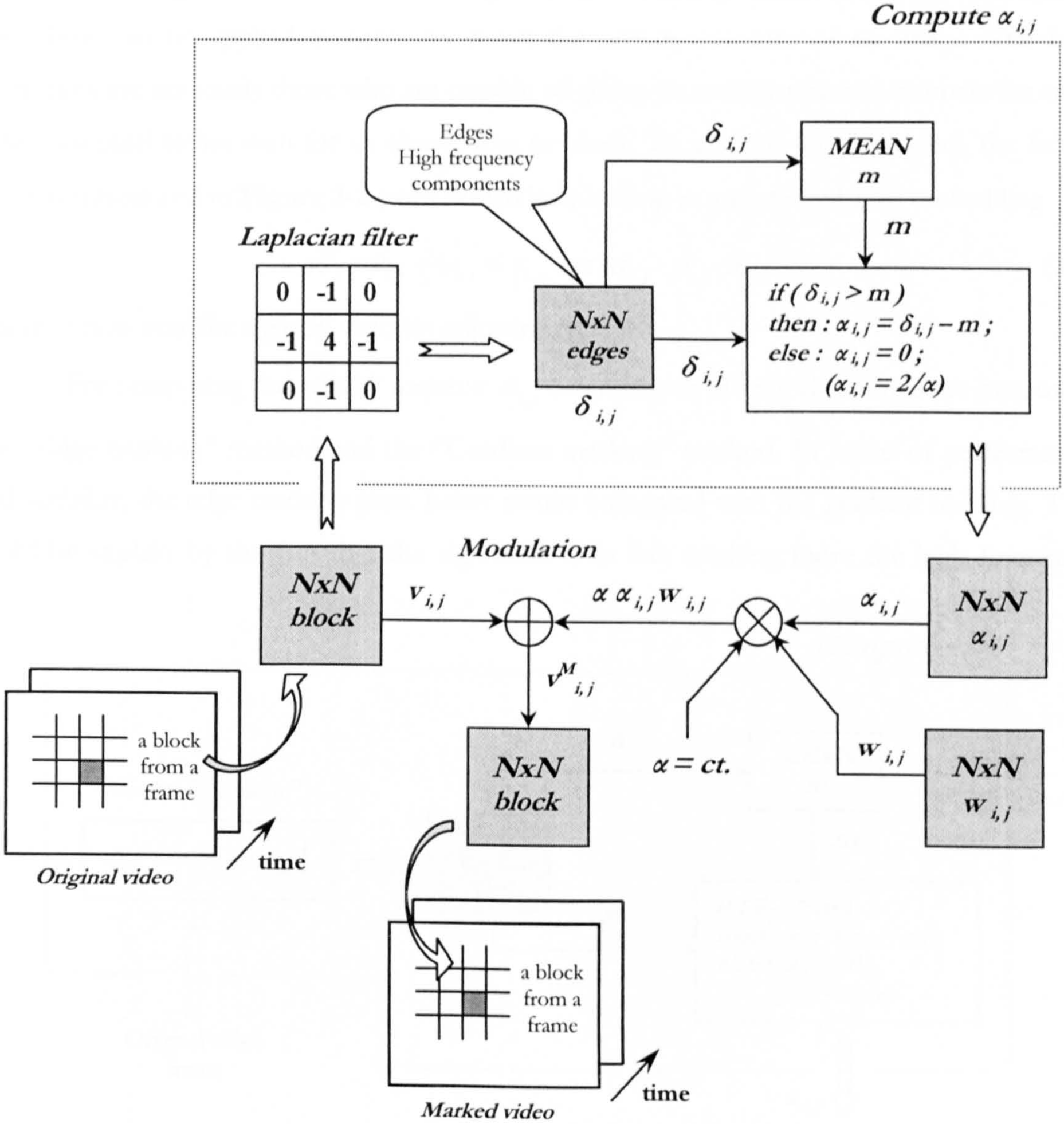


Figure 3-3 Block based "Edge marking" method

3.1.2 Adaptive Watermarking

The embedding algorithm presented in the previous section does not take into account the particularities of the video sequence. Each pixel is marked with the same constant strength, represented by the factor α in equation (3.3).

It is well known from the human vision theory that the eye is more sensitive to changes in the uniform areas of a picture compared with the same changes done in a high activity area which contains various edges and textures. On the other hand dark areas are more susceptible to visibility artefacts than light areas. This makes possible to insert a stronger watermark in the textures and edges while still maintaining the low visibility. Different image processing algorithms can be applied in order to derive the activity measure of an image. The best algorithms are obviously those who are capable of giving an activity measure estimate for each individual pixel rather than for an object, area or block. To account for this aspect, the factor $\alpha_{i,j}$ was introduced in **Figure 3-2** (with dotted line) leading to an activity based embedding

$$v_{i,j}^{M} = v_{i,j} + w_{i,j} = v_{i,j} + \alpha \cdot \alpha_{i,j} \cdot p_{i,j} \cdot b_{i,j} \tag{3.5}$$

where α now acts like a global visibility adjusting factor.

For computing the activity measure $\alpha_{i,j}$ two relatively simple algorithms are proposed: the "Edge marking" method and the "Gradient marking" method. In terms of performance and visibility, the edge marking gives better results compared with the gradient marking. This could be explain by the fact that the algorithm is in fact marking more the high frequency

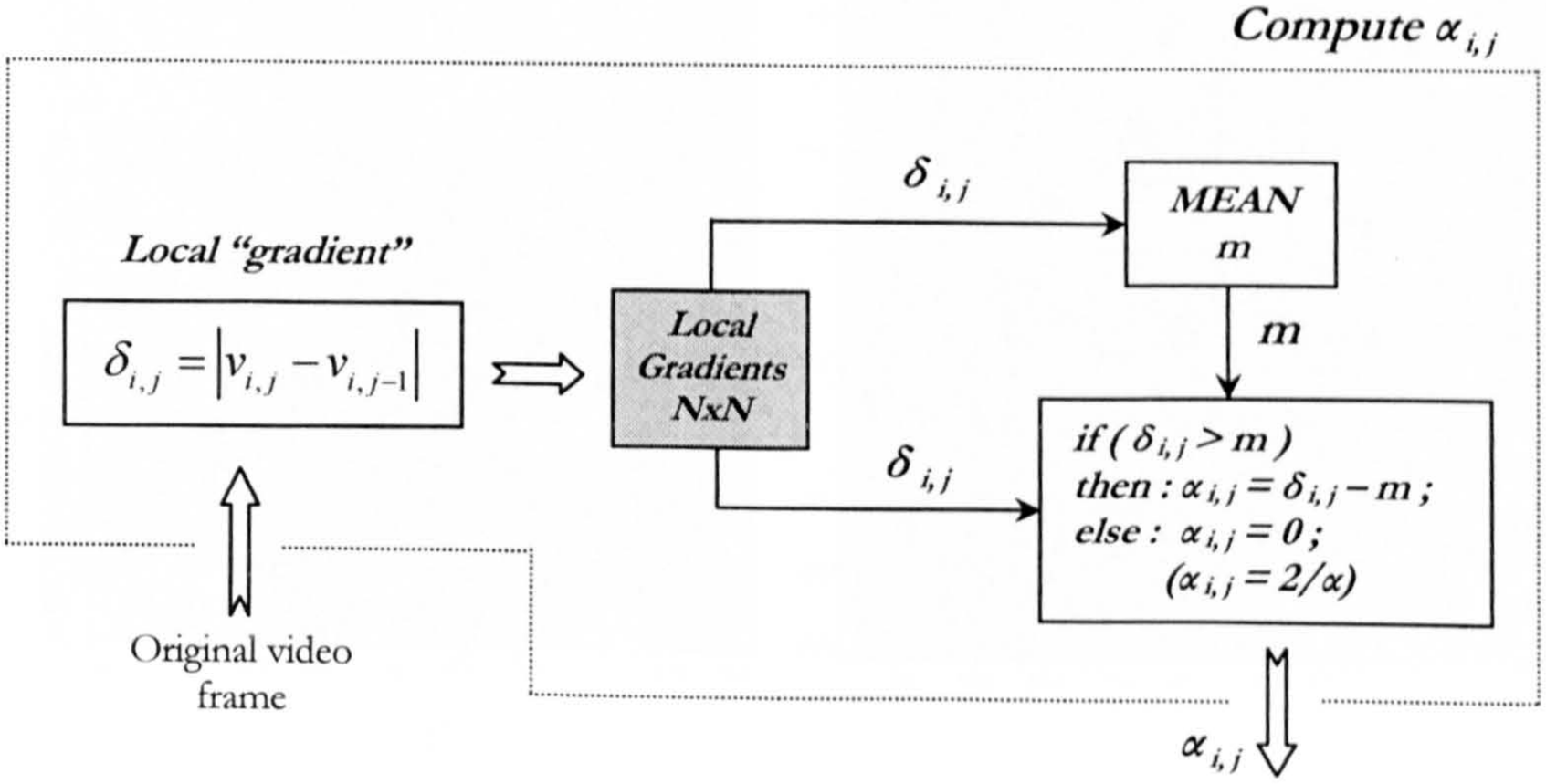


Figure 3-4 Block based "Gradient marking" method

components of the video, due to the type of filter chosen as an activity measure.

Figure 3-3 shows the block scheme of the marking algorithm. The activity measure of the video is given by a high pass filter, in fact a Laplacian filter which acts as a non-directional edge detector. The values returned by the filter are first translated to positive integer values in the range 0…255 (not shown in the figure) and then processed following the algorithm described in the figure. The purpose of this operation is to derive an activity measure which assigns a higher $\alpha_{i,j}$ value to the highly textured areas and to the edges, while keeping $\alpha_{i,j}$ low for uniform areas which are very sensitive to noise. As **Figure 3-3** illustrates, this is done by computing the mean of the filtered values which is used both as a threshold and as limiting factor when a given value exceeds it. This limits the amplitude of the watermark inserted in the edges in order to keep the visibility of the mark at low levels. In the low detail regions and

Figure 3-5 Activity marking: **(a)** Original image, **(b)** the result after Laplacian filtering, **(c)** the factor $\alpha_{i,j}$ for "Edge marking" and **(d)** the factor $\alpha_{i,j}$ for "Gradient marking"

uniform regions, the algorithm uses a constant value usually set around a value of 2, which proves to be low enough for most pictures. The global factor α can be used to boost or lower the visibility of the watermark depending on the particular picture. The experiments show that the edge marking performs quite well in terms of visibility; for the same visibility of the watermark the edge marking scheme performs better than a uniform marking scheme.

A similar algorithm is described in **Figure 3-4**. This time instead of convolving the picture with a spatial filter, the activity measure is obtained simply by subtracting two additional pixels. The algorithm used to process the local gradients is identical with the one used in the first case. This method is simpler and quicker, but tends to amplify too much the strong edges and to reduce too much the weak edges. This leads to increased visibility of the watermark around edges and overall gives lower performance. For both methods the maximum value of the factor α is around 1/3.

Figure 3-5 shows a graphical representation of the factor $\alpha_{i,j}$ for the well known image "Lena", using both methods. It is easy to see, when one compares **(c)** with **(d)** that the gradient marking method accentuates too much the strong edges within the image in expense of the moderate edges and textures. This leads to visible artefacts around the significant edges within the image, and as a result the overall amplitude of the watermark has to be reduced, leading to a decrease in robustness.

3.2 Watermark Recovery

3.2.1 Retrieving of the Watermark

For a blind system, watermark retrieval follows basic spread-spectrum receiver theory and uses a matched filter (correlation detector) to extract the watermark bits from the received

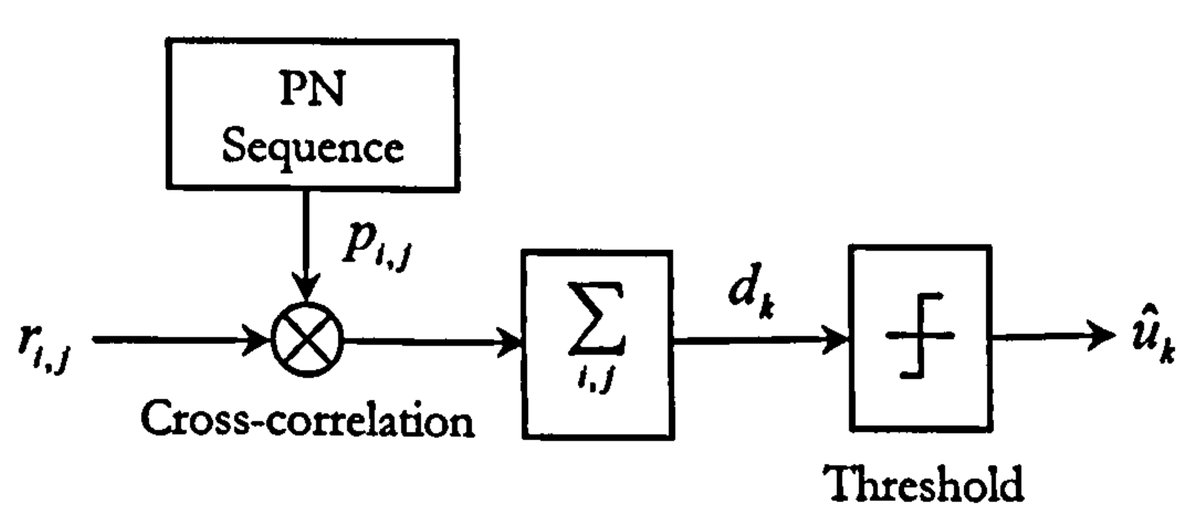


Figure 3-6 Watermark detection

video sequence. Essentially a matched filter is designed to maximise the SNR at a point in time. It does not preserve the input signal wave shape. This is not the objective of a matched filter. The objective is to distort the input signal wave shape and filter the noise so that at the sampling time the output signal level will be as large as possible with respect to the output noise level [Couch, 1987]. A particularly popular form of a matched filter is the correlation detector, often used for band pass signals. One example could be the detection of BPSK (Binary Phase Shift Keying) signals.

General schematic of the watermark detection process is illustrated in **Figure 3-6**.

Each data bit is extracted by cross-correlation of received sequence $r_{i,j}$ with the same PN sequence $p_{i,j}$ that was used for embedding, over a window of c_r bits. Assuming that the watermarked video was not attacked in any way, and therefore $r_{i,j} = v^M_{i,j}$, the detection process can be described as

$$\begin{aligned} d_k &= \sum_{(i,j)=kc_r}^{(k+1)c_r-1} p_{i,j} r_{i,j} \\ &= \sum_{(i,j)=kc_r}^{(k+1)c_r-1} p_{i,j} v^M_{i,j} \\ &= \sum_{(i,j)=kc_r}^{(k+1)c_r-1} p_{i,j} v_{i,j} + \sum_{(i,j)=kc_r}^{(k+1)c_r-1} \alpha b_{i,j} p^2_{i,j} \end{aligned} \tag{3.6}$$

Assuming that the PN sequence $p_{i,j}$ and the original video sequence $v_{i,j}$ are uncorrelated over a large window c_r the first term from the equation (3.6) is close to zero and the equation (3.6) can be rewritten as

$$d_k \approx \sum_{(i,j)=kc_r}^{(k+1)c_r-1} \alpha b_{i,j} p^2_{i,j} \tag{3.7}$$

As $p^2_{i,j} = 1$ for a binary PN sequence and because over the window c_r the spread bits $b_{i,j}$ simply take on the value of the data bit u_k, the equation (3.7) becomes

$$d_k \approx \alpha c_r u_k \tag{3.8}$$

and u_k can be detected as

$$\hat{u}_k = sign(d_k) \tag{3.9}$$

by using a simple threshold.

However, in practice the first sum from equation (3.6) is not quite zero. Hartung suggests using a correction factor Δ which accounts for the different number of 1's and -1's in the PN sequence

$$\Delta = -\left(\sum_{(i,j)=kc_r}^{(k+1)c_r-1} p_{i,j}\right) mean\left(r_{i,j}\right) \tag{3.10}$$

As experimental results show, this correction leads only to marginal improvements (section 3.3).

The technique described above works very well only when the correct synchronisation between the PN sequence $p_{i,j}$ and the marked video sequence $r_{i,j}$ is maintained. Once an attack affects the synchronisation (the simplest way to "achieve" that is to cut a line/column or to slightly shift the video frame for example) the scheme is incapable of retrieving the watermark. This is a typical weakness of any spread-spectrum technique and obviously constitutes a major handicap for any robust watermarking scheme. In order to overcome this major problem one has to use a "sliding correlator" which effectively searches for the right peak or in other words re-establishes the correct synchronisation between the received sequence and the PN sequence.

The noise at the receiver, represented by the original video has much higher amplitude compared with the watermark itself and this is heavily affecting the cross correlation process. Since this component is not necessary for detection, it's a good idea to remove it. In fact, in communication systems is customary to use a pre-filter prior to cross-correlation as this improves the performance of the receiver. Following this reasoning, a pre-filter is employed prior to the cross-correlation process. Practically the received signal $r_{i,j}$ is high pass filtered in order to separate and remove the major components of the video signal itself. The filter used is a simple 3x3 spatial filter, with a Laplacian kernel. The Laplacian filters are characterised by the fact that the sum of their kernel coefficients is always zero which is equivalent to say that this filter rejects the DC component of the input signal. Therefore the filter will cut the low frequencies and the DC component of the video sequence, breaking the strong dependence between the cross-correlation process and these components. As the results will show (section 3.3.1), the filtering increases the performance of the system substantially.

3.2.2 Sliding Window Cross-Correlators

The importance of a sliding window correlator was emphasised in the previous section. Indeed one has to counteract the effects of the de-synchronisation attacks in order to have a

robust system. At this time, only the case of a two dimensional (2-D) cross-correlator is considered. An improved version capable to tackle the 3-D case will be presented in Chapter 5.

The main problem of a sliding-window correlator is its complexity. Indeed even for the 2-D case the searching space grows very quickly with the "size" of the window. In other words, the sliding window has to move in a 2-D searching area defined by two parameters: a horizontal offset and a vertical offset specifying how much the window will slide left and right

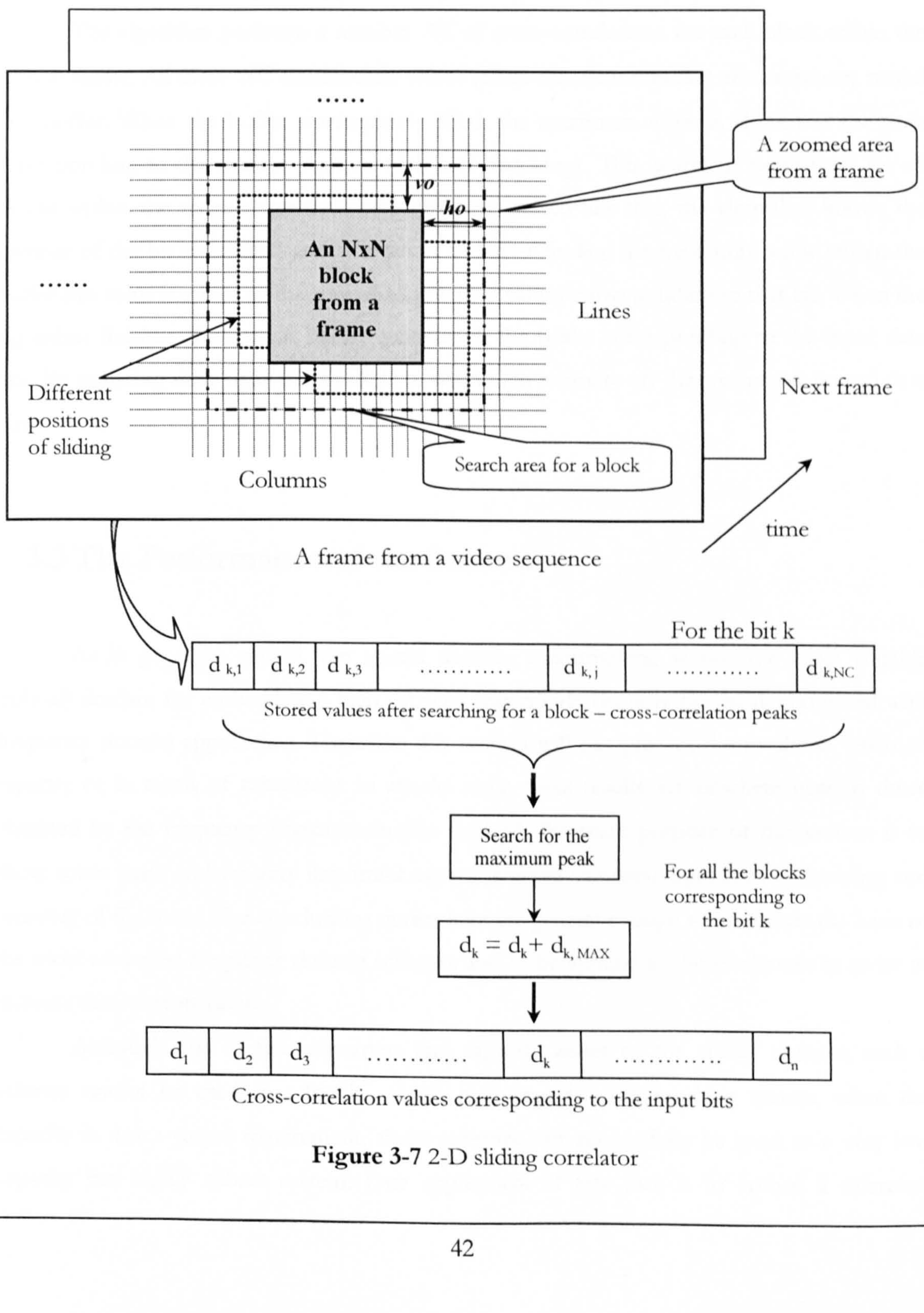


Figure 3-7 2-D sliding correlator

and respectively up and down of a reference point given by the position where the original block is expected to be, for both horizontal and respectively vertical locations.

Therefore the number of the cross-correlations which have to be performed can be defined as

$$NC = (2 \cdot ho + 1)(2 \cdot vo + 1) \tag{3.11}$$

where ho and vo represent the horizontal and respectively the vertical offsets, as described above. The schematic of the entire process is presented in **Figure 3-7.**

The algorithm performs a number NC of cross-correlations for each block within the current frame. All these NC intermediate values (cross-correlation peaks) are temporary stored in a buffer. When the buffer is completely filled, the maximum value is chosen as the peak corresponding to the correct position (the best matching). This search is performed for all blocks within the video sequence. For each block parsed like this, the algorithm knows the number of the bit embedded at that location and after finding the maximum value within the buffer this value is added to the sum of all previous values corresponding to that bit. When the algorithm finishes, the second buffer contains all the peaks corresponding to the input data bits. By analysing their signs it is possible to obtain (an estimate of) the watermark (input) data bits.

3.3 The Performance of the Scheme

As is generally agreed, the spatial domain watermarking is not the most suitable (robust) domain for embedding a high capacity watermark, being rather weak compared with frequency domain approaches. Therefore this section will not present the results in terms of capacity or in terms of robustness to attacks since these results are nowhere near to those obtained by the frequency domain schemes. Instead, the main purpose of this section is to show some basic and yet very important aspects relative to spread spectrum embedding and recovery of the mark. The conclusions drawn here are general enough to constitute the basis of the most advanced frequency domain schemes and to be applied to these schemes in order to increase their performance.

Although due to the robustness and capacity issues of the spatial domain, such a scheme cannot be used as a highly robust, high capacity watermarking system, when the capacity is not a major requirement, these schemes can successfully be used as a very low capacity but highly robust system. One application of this case is to embed a reference

watermark (1 bit watermark). This case is described in detail in Chapter 7 and its performance is analysed for a wide range of attacks.

This section will analyse several aspects like the effect of high pass filtering prior to cross-correlation, the effect of various block dimensions for both embedding and retrieving, and finally, the effect of 2-D sliding on the performance of the system. These results are presented and compared for both uniform marking and edge marking schemes. Most of these findings will constitute the basis for the more advanced DCT and DWT systems presented in Chapter 5 and respectively Chapter 6.

As **Figure 3-8** clearly indicates, pre-filtering drastically improves the performance of the system. Assuming that no sliding is performed – and this is the case illustrated in **(b)** and **(d)** – filtering improves the SNR up to 10 times for uniform marking and up to 6 times for edge marking. When sliding is performed – as shown in **Figure 3-8 (a)** and **(c)** – than the gain is up to 8 times for uniform marking and up 14 times for edge marking.

All the results presented in this section are for 25 frames (1 second) of "basketball" video sequence, for a length of the watermark of 64 bits. The amplitude factor α was set to 2 for uniform marking and to 0.2 for edge marking, which leads to a low visibility watermark. The vertical axis in **Figure 3-8**, **Figure 3-9**, **Figure 3-10** and **Figure 3-11** represents the SNR

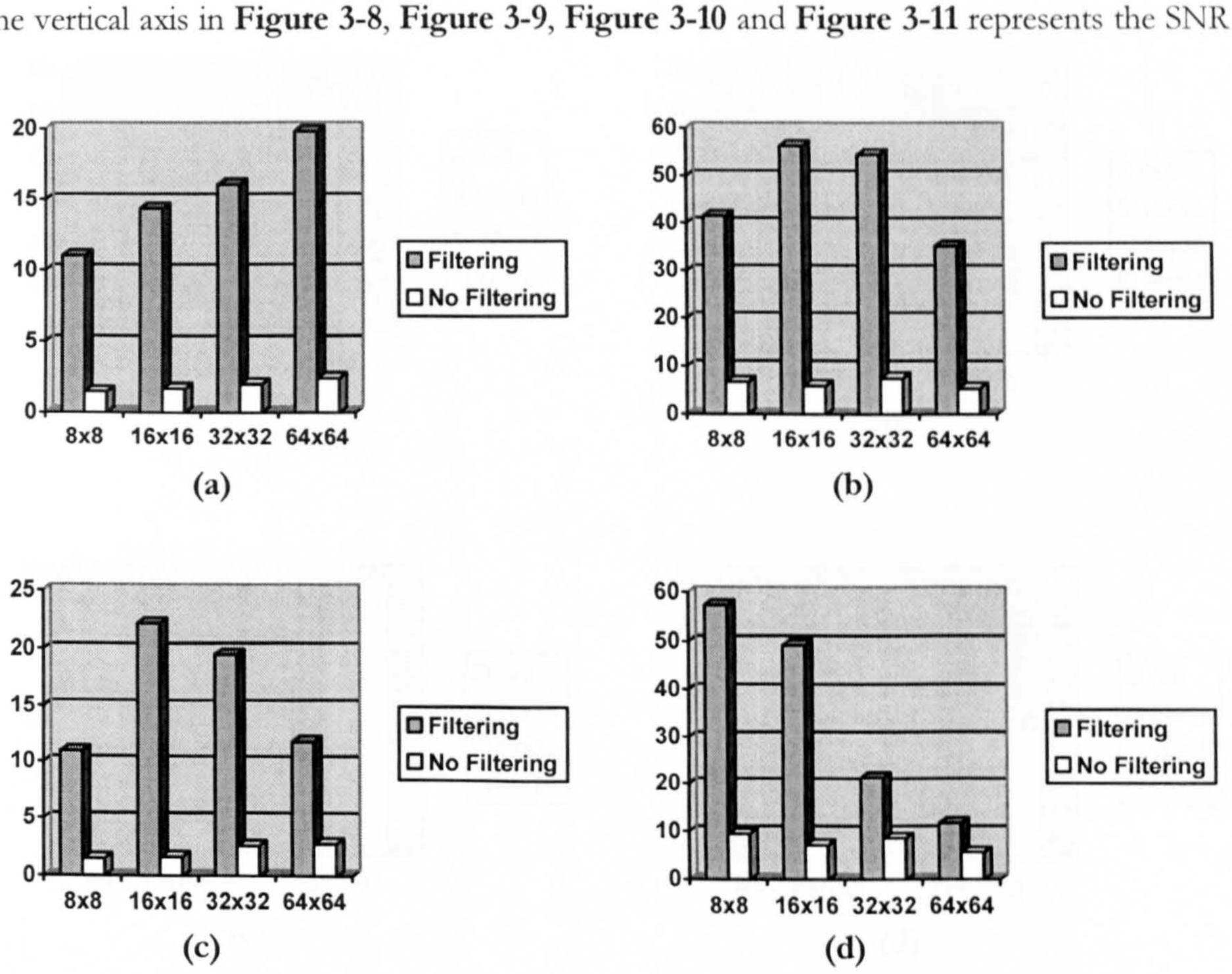


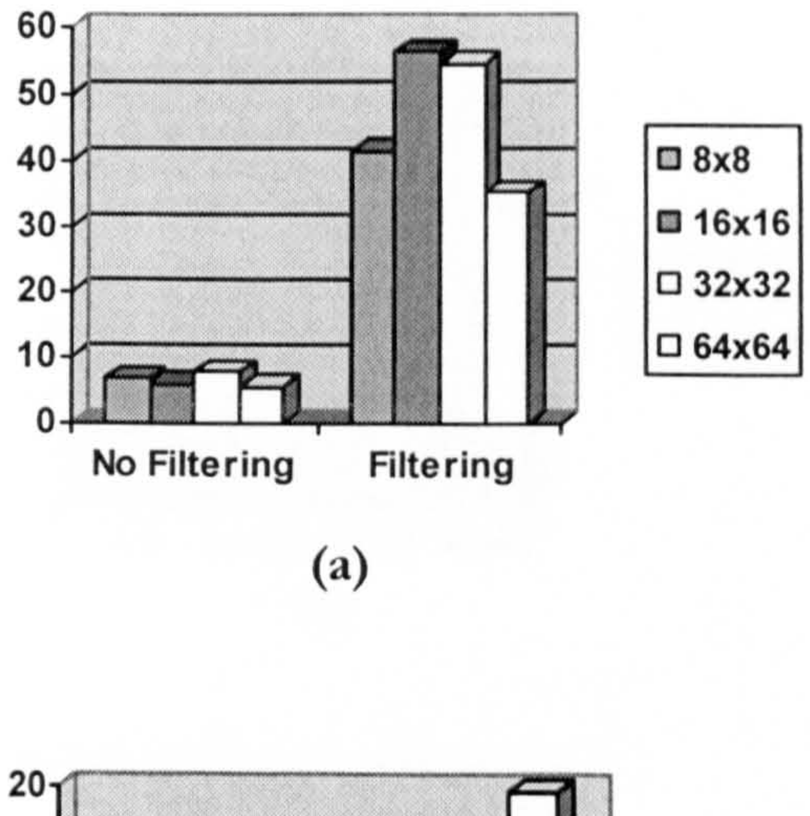
Figure 3-8 The effect of filtering for: **(a)** Uniform marking, 2x2 sliding, **(b)** Uniform marking, no sliding, **(c)** Edge marking, 2x2 sliding and **(d)** Edge marking, no sliding.

ratio of the cross-correlation peaks.

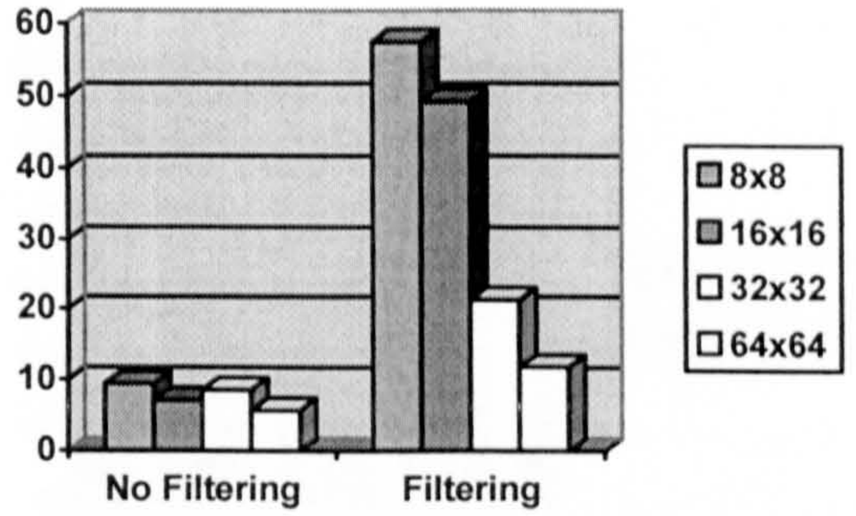
One advantage of using either 8x8 or 16x16 blocks is that 8 and 16 divides exactly the dimension of the frame which for ITU-R 601 video signals is 720x576. This is very convenient for implementation. Unfortunately this is not the case for 32x32 or 64x64. So rather than complicating the implementation the program will mark the nearest smaller area of the frame which in this case is 704x576. Therefore 16 lines are "lost", e.g. not marked. This reduces slightly the chip rate from 162000 to 158400. Anyway the difference in performance because of those 16 lines is not big enough to justify the effort of implementation. This is suggested by the results presented in **Figure 3-9 (a)**. It can be seen that the difference between the 16x16 and 32x32 cases is quite small.

Block size	8x8	16x16	32x32	64x64
Number of blocks per frame	6480	1620	396	99

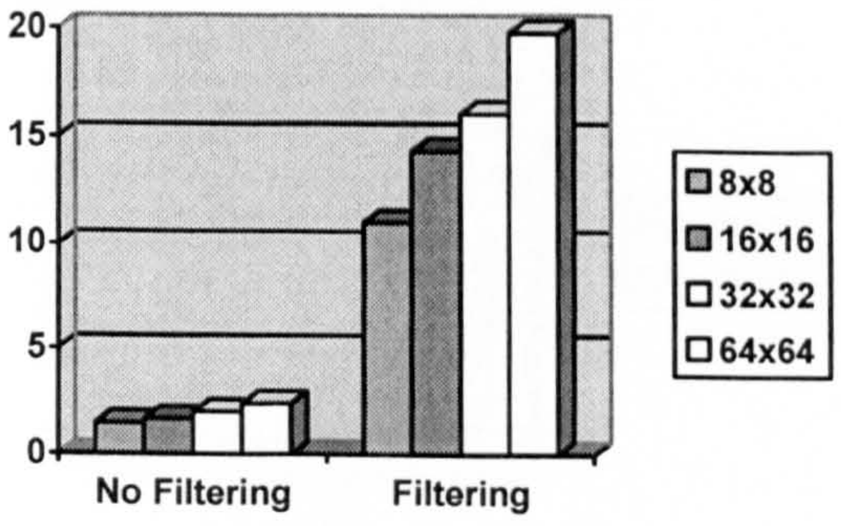
Table 3-1 The number of blocks per frame for several block sizes.



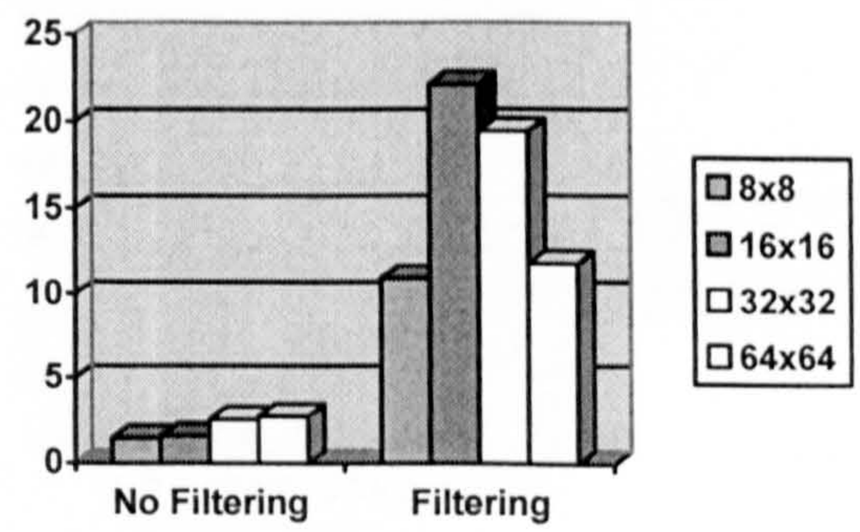
(a)



(b)



(c)



(d)

Figure 3-9 The effect of block size for: **(a)** Uniform marking, no sliding, **(b)** Edge marking, no sliding, **(c)** Uniform marking, 2x2 sliding and **(d)** Edge marking, 2x2 sliding.

Ignoring the non-filtering cases, it can be seen that the edge marking scheme performs increasingly worse as the block size increases - **Figure 3-9 (b)**. This is due to the visual model used which for higher block sizes is less and less adaptive. This suggests using an 8x8 block dimension for computing the visual model, even if the actual marking and recovery is done on a different block size. This approach was successfully followed for the DCT-based watermarking system presented in Chapter 5.

Another important aspect of choosing the block size can be observed in **Table 3-1**. The number of blocks per frame drops significantly (by a rate of 4 in this case) with the block size. This fact leads to only 99 blocks per frame for a block size of 64x64. This is fine for a relatively low number of data bits (<99), but taking into account that an entire block corresponds to only one data bit, if the length of the watermark is higher than 99 then there is not enough room in one frame to embed all the data bits. This complicates even more the embedding, could reduce the efficiency of the visual model, reduces the security of the algorithm and has serious robustness implications. For example in the case of line cuts, or column cuts, and for cropping, smaller block sizes lead to better performance.

The main advantage of choosing a higher block size is performance under sliding. As a

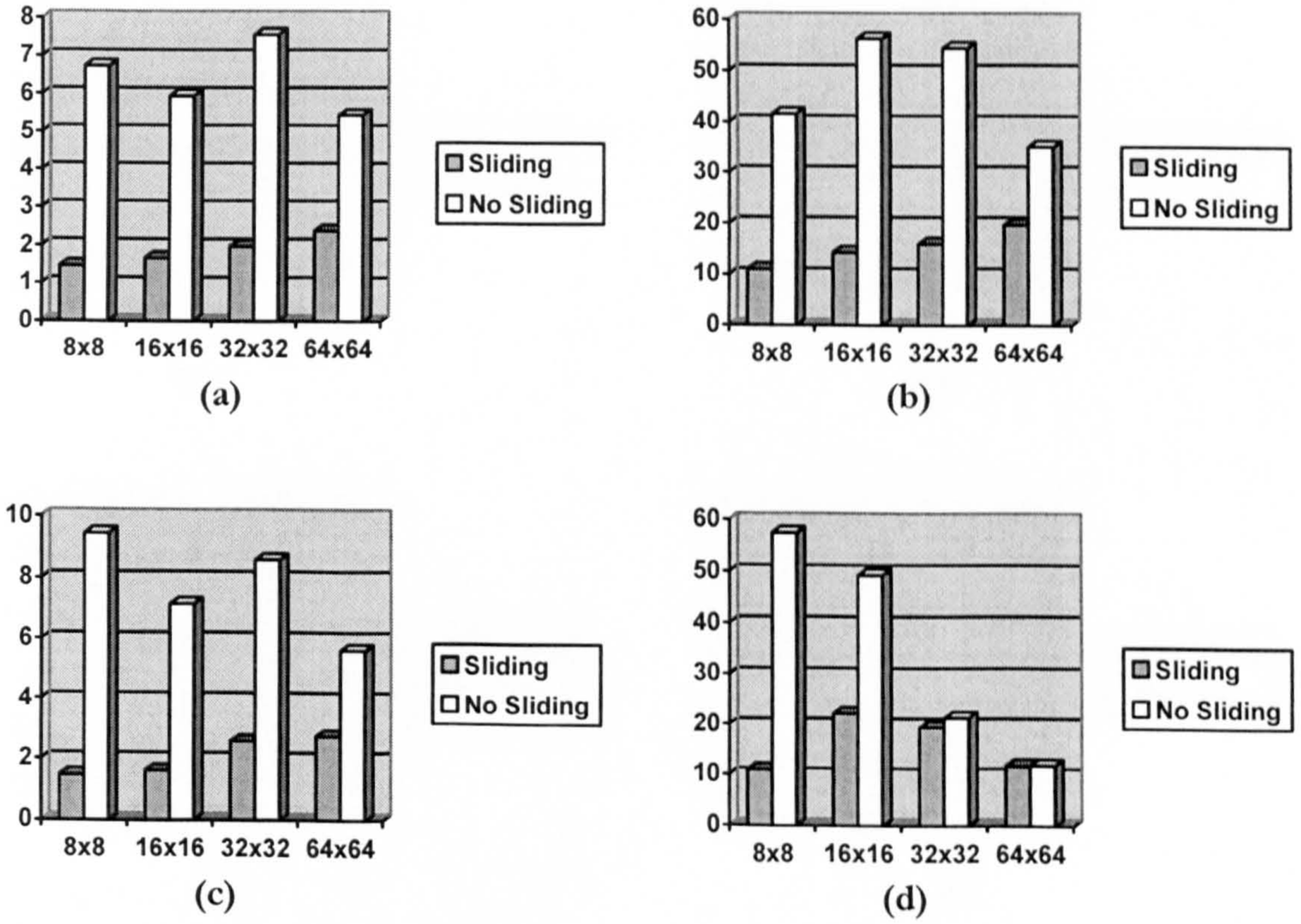


Figure 3-10 The effect of sliding for: **(a)** Uniform marking, no filtering, **(b)** Uniform marking, with filtering, **(c)** Edge marking, no filtering and **(d)** Edge marking, with filtering.

rule, the higher the block size the better the performance when a sliding correlator is used. This can be seen in **Figure 3-9 (c)**. For example just by using a block size of 16x16 instead of 8x8 the SNR increases with 27%. For a block size of 32x32 and 64x64 this percentage increases to 45% and respectively 79%. To resolve these conflicting aspects, one has to choose a compromise solution. The experiments show that a block size of 16x16 is a reasonable choice.

The effects of the sliding compared with the no-sliding situation are illustrated in **Figure 3-10**. The marked sequence wasn't attacked in any way, so these results suggest that the sliding correlator has an inherent loss even when the marked video was not attacked. Chapter 5 will show this aspect in more detail, together with the results of the sliding correlator for line/column cuts. This loss could be quite high as **Figure 3-10** shows. When no filtering is performed the loss can be up to 6 times for uniform marking and up to 5 times for edge marking. With filtering the loss is reduced to up to 4 times for uniform marking and remains approximately the same (5 times) for edge marking. **Figure 3-10** shows once more that the higher the block size is, the better is the performance which can be expected. Moreover, the difference between the sliding and non-sliding cases decreases as the block size increases.

Finally, the effect of average compensation is illustrated in **Figure 3-11**. As stated at the

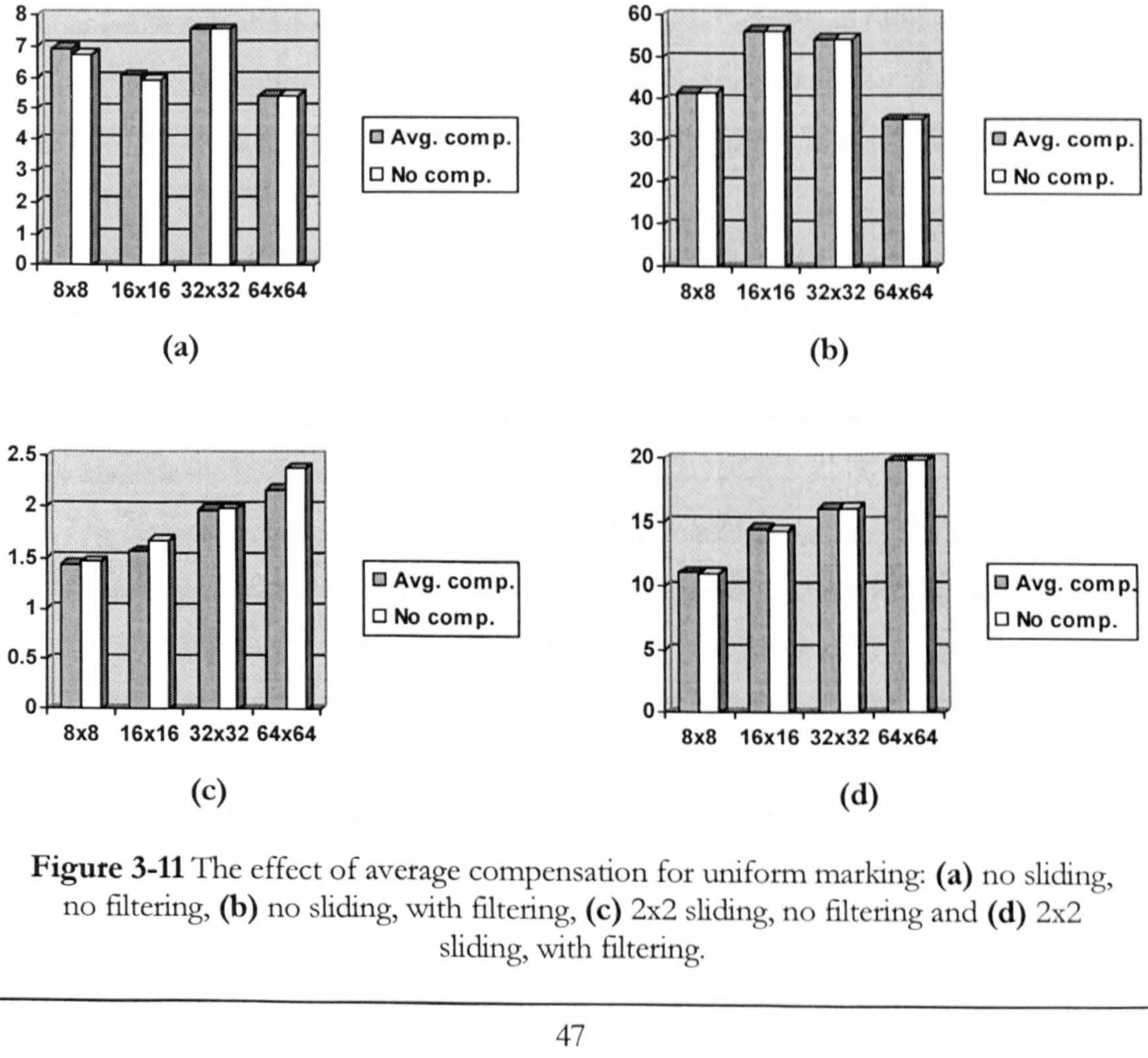


Figure 3-11 The effect of average compensation for uniform marking: **(a)** no sliding, no filtering, **(b)** no sliding, with filtering, **(c)** 2x2 sliding, no filtering and **(d)** 2x2 sliding, with filtering.

end of section 3.2.1, the average compensation doesn't really make a difference; in some instances the performance is even a little worse, although generally speaking average compensation improves very slightly the SNR. This improvement is so small that is not justifying its implementation in a practical scheme.

3.4 Conclusions

The results presented in the previous section show that a block based approach is preferred for a robust multi-bit watermarking system from several reasons. Robustness to attacks is perhaps the most important reason. Another advantage is that the HVS models tend to work better for smaller blocks. Finally, the security of the system is improved and overall the system is more flexible.

The HVS models improve the watermark invisibility and the robustness of the scheme. This is happening because the HVS model embeds more energy into the video sequence and therefore the SNR of the cross-correlation peaks will be higher.

The recovery of the watermark is greatly improved by pre-filtering the video sequence with a Laplacian kernel prior to cross-correlation. A simple 3x3 spatial filter improves the SNR of the cross-correlation peaks up to 10 times. The simulations show that the correction factor Δ suggested by Hartung is not effective, leading only to marginal improvements and therefore its presence is not justified.

In order to resynchronise a geometrically attacked video signal, a 2-D sliding correlator is needed. The downside of employing a sliding window correlator is its complexity; therefore sliding is a time consuming operation, difficult to implement in a real time system. Moreover, the sliding correlator has an inherent loss even in the case of un-attacked video; by using a sliding correlator the SNR can drop up to 6 times.

The dimension (size) of the block is a very important factor. The HVS model performs better for smaller blocks (8x8), while the recovery of the watermark works better for higher block sizes. The higher the block size, the better the performance of the system under sliding but higher block sizes are worse in terms of system's robustness, security and complexity and therefore a compromise has to be reached. As a compromise, a block size of 16x16 gives acceptable results. In conclusion it is better to use a mixed approach (Chapter 5), where marking is performed on smaller blocks (8x8) while the recovery is performed on higher block sizes (e.g. 16x16).

Chapter 4

"Although the rivers and mountains of the world have not changed, their ancient and modern names are different."

Wen Zhuang, Ming dynasty

Channel Capacity & Turbo Coding

This chapter presents some basic communication principles related to channel capacity and forward error correction (FEC) codes. The Shannon's channel capacity theorem and its practical implications are also discussed, as well as the ways of getting closer to this limit. It is shown that in order to achieve better performance in a communication system, e.g. to get closer to the Shannon's limit, one can use the new state-of-the-art FEC codes. One of the best error correcting codes available today are the Turbo codes. A short introduction to Turbo codes and their characteristics and performance is also provided in this chapter.

The watermarking is seen by the information theory perspective and therefore by applying this theory and by using Turbo coding, the performance of the watermarking system is greatly improved. This conclusion can be easily drawn by analysing the results presented in Chapter 5 and Chapter 6.

4.1 The Channel's Capacity

Broadband providers are naturally interested in increasing transmission distance and data rates, on the one hand to cut down on the amount of physical plant that must be installed, and on the other to increase throughput. To accomplish this, the broadband provider may be tempted to extend the antenna length or increase the transmission power, but these are costly and oftentimes unacceptable alternatives.

An alternative that is becoming more attractive for providing increased performance is the use of powerful FEC. Embedding an FEC codec that implements state-of-the-art coding technology in the transceiver can radically increase the transmitted-data rate or transmission distance, or alternatively decrease the required antenna size and power.

FEC is the addition of redundancy (e.g., parity-check symbols) to a transmitted message, allowing the receiver to decode the received message, check symbols, and correct some limited number of errors in the received-data stream. The ability of FEC to increase the signal-to-noise capability of a communications channel depends on the code used and the channel characteristics.

All channels have a theoretical limit for information rate content at a constant signal-to-noise ratio (SNR) known as the Shannon capacity. The Shannon capacity limit defines the maximum information content for any particular channel. Communications systems that do not use FEC operate far from this limit, often 10 dB or more. Examples of these systems include voice applications and other communications systems where an occasional bit error can be tolerated. To achieve the accuracy and data rates required most of the time (wireless Internet access for example), the system without FEC would require 10 dB greater SNR than a "perfect" system operating at the Shannon capacity.

The use of traditional FEC codes such as Reed-Solomon (RS) coding substantially improves the efficiency of the communications channel allowing operation much closer to the Shannon capacity. For a typical channel, the addition of RS coding allows the system to operate within approximately 4 dB of the Shannon capacity (depending on channel characteristics). The resulting benefit translates into higher data rates, lower bit-error rates (BER), greater transmission distance and greater immunity to interference effects. However, this still leaves considerable room for improvement. After all, to make up for the 4-dB distance from optimum, the system developer must spend valuable resources in terms of transmission power, antenna size and bandwidth.

As an example, a more powerful code that provides a 3-dB coding gain over the RS coding can mean a reduction in antenna diameter by 30 percent, a decrease in transmission power by a factor of two, a transmission distance increased by 40 percent, or increased data throughput by a factor of two. Recent breakthroughs in error-correction coding have led to new FEC codes that can provide this 3-dB performance gain over RS coding. The Turbo codes are one possible example. In some special circumstances, it is possible to approach the Shannon's capacity limit by 0.27 dB using Hamming codes in a turbo decoding scheme [Nickl et al, 1997].

4.1.1 The Noisy Channel Coding Theorem

In 1948, Shannon derived the following formula for the capacity of an additive white Gaussian noise channel (AWGN)

$$C = W \log_2\left(1+\frac{S}{N}\right) \tag{4.1}$$

where the capacity is expressed in bits/sec, W represents the bandwidth of the channel, S is the average signal power and N is the total average noise power of the channel.

Shannon established the noisy channel coding theorem:

1. For information rate $R_{info} < C$ there exists a coding system with arbitrarily low block and bit error rates as we let the code length $n \to \infty$.

2. For information rate $R_{info} > C$ the bit and block error rates are strictly bounded away from zero for any coding system.

The noisy channel coding theorem therefore establishes rigid limits on the maximal supportable transmission rate of an AWGN channel in terms of power and bandwidth.

To characterise how efficiently a system uses its allotted bandwidth, one can define the *bandwidth efficiency* as

$$\eta = \frac{C}{W} \tag{4.2}$$

The Shannon limit can be calculated as

$$\eta_{\max} = \log_2\left(1+\frac{S}{N}\right) \tag{4.3}$$

Taking into account that

$$S = \frac{kE_b}{T} = RE_b \tag{4.4}$$

where E_b represents the energy per bit, k is the number of bits transmitted per symbol, T is the duration of a symbol and R is the transmission rate (code rate) of the system. Now the Shannon limit can be obtained in terms of the bit energy and noise power spectral density

$$\eta_{\max} = \log_2\left(1+\frac{RE_b}{WN_0}\right) \tag{4.5}$$

where $N = WN_0$ represents the total noise power and N_0 is the one-sided noise power spectral density.

The equation (4.5) can be resolved in order to obtain the minimum bit energy required for reliable transmission, e.g. the *Shannon bound*

$$\frac{E_b}{N_0} \geq \frac{2^{\eta_{\max}} - 1}{\eta_{\max}} \tag{4.6}$$

From equation (4.6) it is possible to establish the fundamental limit for reliable communication. This can be obtained by considering an infinite amount of bandwidth, i.e. $\eta_{\max} \to 0$

$$\frac{E_b}{N_0} \geq \lim_{\eta_{\max} \to 0} \frac{2^{\eta_{\max}} - 1}{\eta_{\max}} = \ln(2) = -1.59dB \tag{4.7}$$

This represents the absolute minimum signal energy to noise power spectral density ratio required to reliably transmit one bit of information, even for unlimited bandwidth or bit rate tending to zero.

The dependence on the arbitrary definition of the bandwidth W is usually not satisfactory. The answer is to normalise these formulas per signal dimension [Wozencraft et al, 1965]. This is useful when the question of waveforms and pulse shaping is not a central issue, since it allows one to eliminate these considerations by treating signal dimensions. In this case Shannon's capacity and the corresponding bound are

$$\begin{aligned} C_d &= \frac{1}{2}\log_2\left(1 + 2\frac{R_d E_b}{N_0}\right) \\ \frac{E_b}{N_0} &\geq \frac{2^{2C_d} - 1}{2C_d} \end{aligned} \tag{4.8}$$

The dependence of Shannon's capacity limit of the code rate is illustrated in **Table 4-1**, for an AWGN channel with QPSK (Quadrature Phase Shift Keying) modulation [Dolinar et al, 1998].

Code rate, R	Capacity limit, $E_b / N_0 \,[dB]$
1/2	0
1/3	-0.55
1/4	-0.82
1/6	-1.08
0	-1.59

Table 4-1 Shannon limit for different code rates

4.1.2 Hardware and Software Decoding

A typical communication system can be represented as in [Barbulescu et al, 1996]. Regardless of its source, the information to be transmitted must be translated into a set of signals optimised for the channel over which we want to send it. As **Figure 4-1** shows, the first step is to use a source encoder block for eliminating the redundant part of the signal in order to maximise the information transmission rate. To ensure the secrecy of the transmitted information one could use an encryption scheme. The most important part of the system in the case analysed here, is to protect the signal against the perturbations introduced by the communication channel, which could lead to errors in the transmitted message at the receiving end. This protection is achieved by FEC, using error correction codes that are able to correct the errors at the receiving end. Finally the modulator block generates a signal suitable for the transmission channel.

The importance of using powerful error correction and the economical and practical benefits of such codes was already underlined in the introduction. From coding theory it is known that either by reducing the data rate or increasing the codeword length or the encoding memory, greater protection or coding gain, can be achieved. Unfortunately at the same time the complexity of typical decoding algorithm such as the maximum likelihood decoding algorithms increases exponentially with the encoder memory and the algorithms become difficult to implement. Therefore the increased error correction capability of long codes requires a very high computational effort at the decoder.

In simple systems, the demodulator block from **Figure 4-1** makes a hard decision of the received symbol and passes it to the error control decoder block. In other words, the demodulator decides which of two logical values 0 or 1 was transmitted. No information is

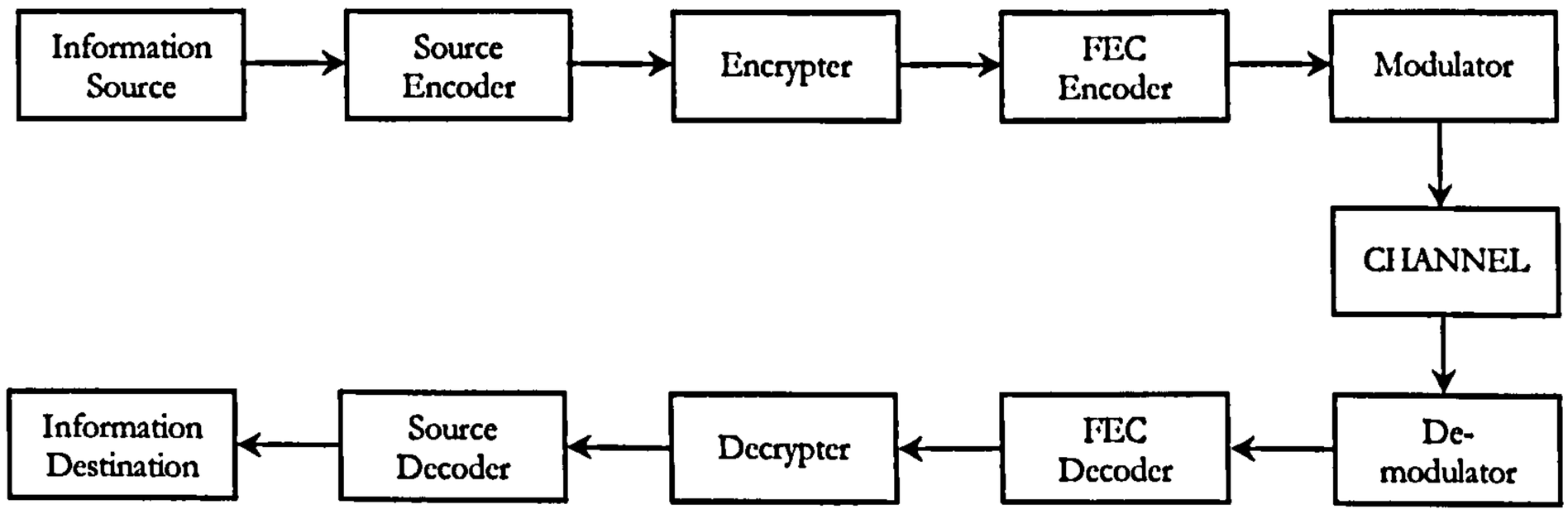


Figure 4-1 The block diagram of a communication system

passed to the FEC decoder about how reliable the hard decision was. Better results can be obtained by using soft input decoding algorithms, e.g. when the quantised analogue received signal is passed directly to the decoder. The same consideration holds for the outputs of the constituent decoders of concatenated codes. By using soft-input-soft-output (SISO) decoders, this information can be passed from one decoder to the next in an iterative fashion. Soft output decision algorithms provide as an output a real number which is a measure of the probability of error in decoding a particular bit. This can be also interpreted as a measure of the reliability of the decoder's hard decision.

It can be shown that the channel capacity of a discrete-input real-output (soft output) memoryless channel (C_{soft}) is greater than that for a discrete-input discrete-output (hard output) memoryless channel (C_{hard}). For a binary symmetric channel with an AWGN distribution and mean value zero this can be proven as follows.

Soft decoding

If the input alphabet is $X=\{x_0,x_1,x_2,...,x_{q-1}\}$ and the output alphabet is $Y=\{-\infty,\infty\}$ then we can define the channel capacity for the soft decoding case as the mutual information between the channel's input and output maximised over all possible channel input distributions $P(x_j)$

$$\begin{aligned} C &= \max_{P(x_j)} I(X;Y) \\ &= \max_{P(x_j)}\left[H(X)-H(X|Y)\right] \\ &= \max_{P(x_j)}\left[H(Y)-H(Y|X)\right] \\ &= \max_{P(x_j)}\sum_{j=0}^{q-1}\int_{-\infty}^{\infty}P(x_j)p(y|x_j)\log\frac{p(y|x_j)}{p(y)}dy \end{aligned} \tag{4.9}$$

Considering the input alphabet restricted to $X=\{-1,+1\}$ and a binary symmetric channel, where $P(-1)=P(+1)=0.5$ then equation (4.9) becomes

$$C_{soft}=\frac{1}{2}\int_{-\infty}^{\infty}p(y|+1)\log_2\frac{p(y|+1)}{p(y)}dy+\frac{1}{2}\int_{-\infty}^{\infty}p(y|-1)\log_2\frac{p(y|-1)}{p(y)}dy \tag{4.10}$$

Taking into account that $p(y)=0.5p(y|+1)+0.5p(y|-1)$ and that for symmetry $p(-y|+1)=p(y|-1)$ then equation (4.10) can be further simplified

$$C_{soft} = \int_{-\infty}^{\infty} p(y|+1)\log_2 \frac{p(y|+1)}{p(y)} dy \tag{4.11}$$

For an AWGN channel with zero mean and variance σ^2 the probability density function $p(y|x=m), m=+/-1$ can be defined as follows

$$p(y|x=m) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right) \tag{4.12}$$

and the variance σ^2 is

$$\sigma^2 = \left(2R\frac{Eb}{No}\right) \tag{4.13}$$

where R represents the code rate and $\frac{Eb}{No}$ is expressed in dB.

Hard decoding

For a hard decoder the output alphabet is finite too so in this case the input and respectively the output alphabets can be defined as $X = \{x_0, x_1, x_2, ..., x_{q-1}\}$ and $Y = \{y_0, y_1, y_2, ..., y_{r-1}\}$. The channel capacity for this case is

$$\begin{aligned} C &= \max_{P(x_j)} I(X;Y) \\ &= \max_{P(x_j)} [H(Y) - H(Y|X)] \\ &= \max_{P(x_j)} \sum_{j=0}^{q-1} \sum_{i=0}^{r-1} P(x_j) P(y_i|x_j) \log \frac{P(y_i|x_j)}{P(y_i)} dy \end{aligned} \tag{4.14}$$

Restricting the input and output alphabets to $X = \{-1, +1\}$ and respectively $Y = \{-1, +1\}$ and taking into account that $P(-1) = P(+1) = 0.5$, for a binary symmetric channel we can define

$$\begin{aligned} P(-1|+1) &= \int_{-\infty}^{0} p(y|+1)\,dy = \hat{P} \\ P(+1|-1) &= \int_{0}^{\infty} p(y|-1)\,dy = \hat{P} \end{aligned} \tag{4.15}$$

Then we can obtain the channel's capacity for the hard decoding case as

$$C_{hard} = 1 + \hat{P}\log_2 \hat{P} + (1-\hat{P})\log_2 (1-\hat{P}) \tag{4.16}$$

For an AWGN channel, $\hat{P}$ can be defined as

$$\hat{P} = Q\left(\sqrt{\frac{2Eb}{No}}\right) \tag{4.17}$$

where the Q function is defined as

$$Q(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}}\left(-\frac{t^2}{2}\right)dt \tag{4.18}$$

Conclusion

The comparison between the soft and hard decoding is illustrated for the binary symmetric channel case in **Figure 4-2**. It can be seen that at low signal to noise ratios C_{soft} is greater than C_{hard} by approximately 2dB. This shows very well the advantage of using soft decision decoding rather then the classical hard decision approach.

4.2 Turbo Codes

Starting with early 1990's concepts like iterative decoding, soft output decision algorithms, special encoding techniques and information transfer techniques were combined in order to create more powerful error correction codes. The combination of these concepts led to appearance of a new class of powerful error correcting codes: the Turbo codes, which made possible communications very close to Shannon's limit. For example, the first Turbo code

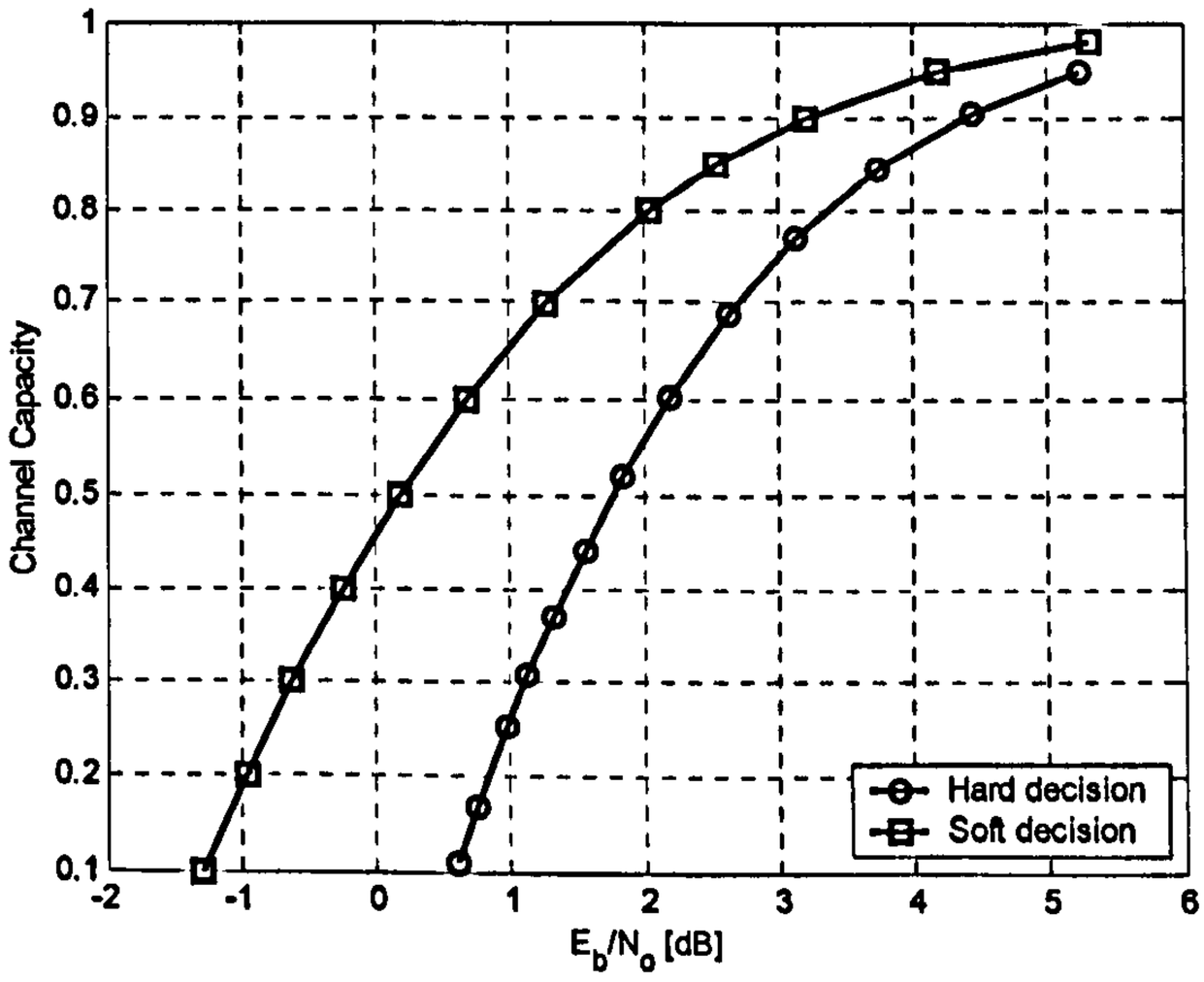


Figure 4-2 Channel capacity for soft and hard decision decoding for a binary symmetric channel

proposed in the literature achieved a bit error rate lower than 10^{-5} within 0.7dB of Shannon's limit.

The Turbo codes were introduced in [Berrou et al, 1993]. They represent a particular class of parallel concatenation of two recursive systematic convolutional codes. In other words, strictly speaking a Turbo code is a 2PCCC (Parallel Concatenated Convolutional Code). Today the term has a more general connotation.

4.2.1 The Structure of a Turbo Code (2PCCC)

The encoder

The block scheme of the encoder is presented in **Figure 4-3**. Since we are dealing with a convolutional code the input sequence is organised in blocks of length N. The first block of data is encoded by the ENC_1 block which is a half rate recursive systematic encoder. The same block of data is interleaved by the interleaver block INT and then encoded by the second encoder ENC_2. Like the first encoder, ENC_2 is a half rate recursive systematic encoder.

The role of the interleaver is to rearrange the order of the information bits from the input. In this way the interleaver increases the minimum distance of the Turbo code and therefore its error correction capability. The design of the interleaver is a key factor which determines the good performance of a Turbo code.

The result is a rate 1/3 turbo code, with the output given by the triplet (v1i, v2i, v3i). Since the code is systematic ui = v1i is the input data at time i and v2i and v3i are the two parity bits at time i. Sometimes the parity bits can be "punctured" using a multiplexing switch in order to obtain higher coding rates.

The decoder

By encoding the same information twice but in different order, the Turbo codes have the advantage of exchanging information between the two constituent decoders. The more "scrambled" the information sequence is for the second encoder the more "uncorrelated" (independent) the information exchange is. This is in fact one of the keys that allows continuous improvement in correction capability when the decoding process is iterated.

As already stated, the Turbo codes use soft output decision algorithms. There are two important categories of soft output decision algorithms. The first category includes the maximum likelihood decoding algorithms which minimise the probability of bit error, such as the maximum a posteriori (MAP) algorithm. The second category includes the maximum

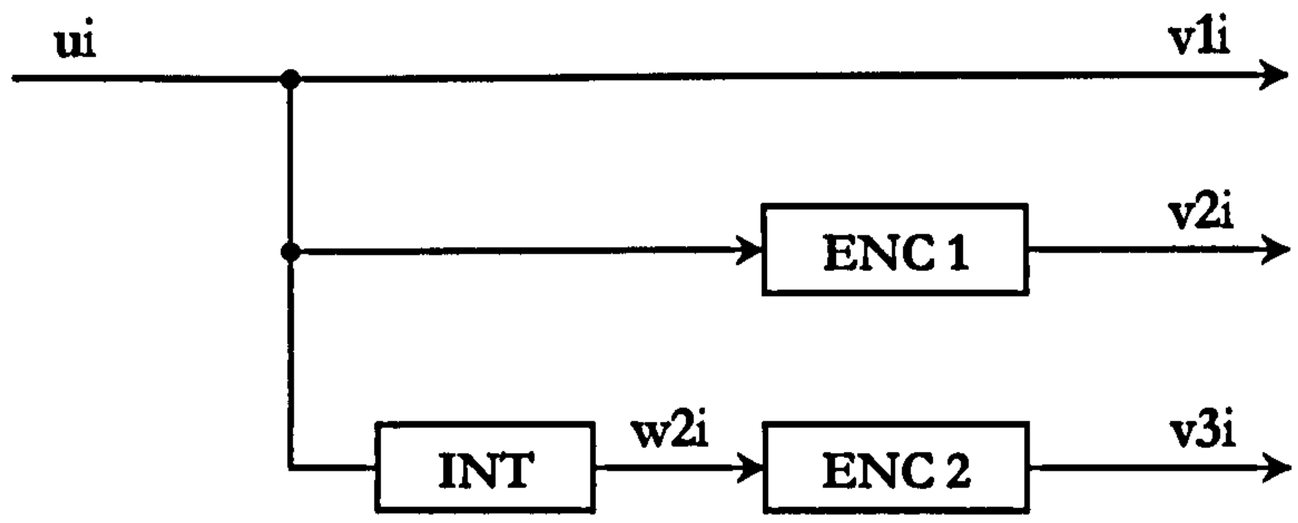


Figure 4-3 The Turbo encoder

likelihood decoding algorithms which minimise the probability of word or sequence error, such as the Soft Output Viterbi Algorithm (SOVA). Although the SOVA algorithm has a soft output, it is sub optimal. The decoder described in this section uses the MAP algorithm and it is presented in **Figure 4-4**.

The decoding principle is briefly described below. First, the MAP2 decoder which corresponds to the encoder ENC2 decodes the information present at its input and initialises the probability Pw20i(r2) with the value 0.5. The decoder MAP2 also calculates the new extrinsic probability Pow20i(r3) using the Pw20i(r2) probability and the Gaussian probabilities of r3i and respectively the interleaved version of r1i, Px20i(r1).

The extrinsic information computed at this step Pow20i(r3), gives a more precise information about the bit w2i. Since w2i is the common information between the two encoders ENC1 and ENC2, the extrinsic information refers to this particular bit. The extrinsic information is then deinterleaved and passed to the first decoder MAP1.

This decoder will start decoding the information from its input, taking into account the extrinsic information supplied by the MAP2 decoder. Therefore when MAP1 decoder

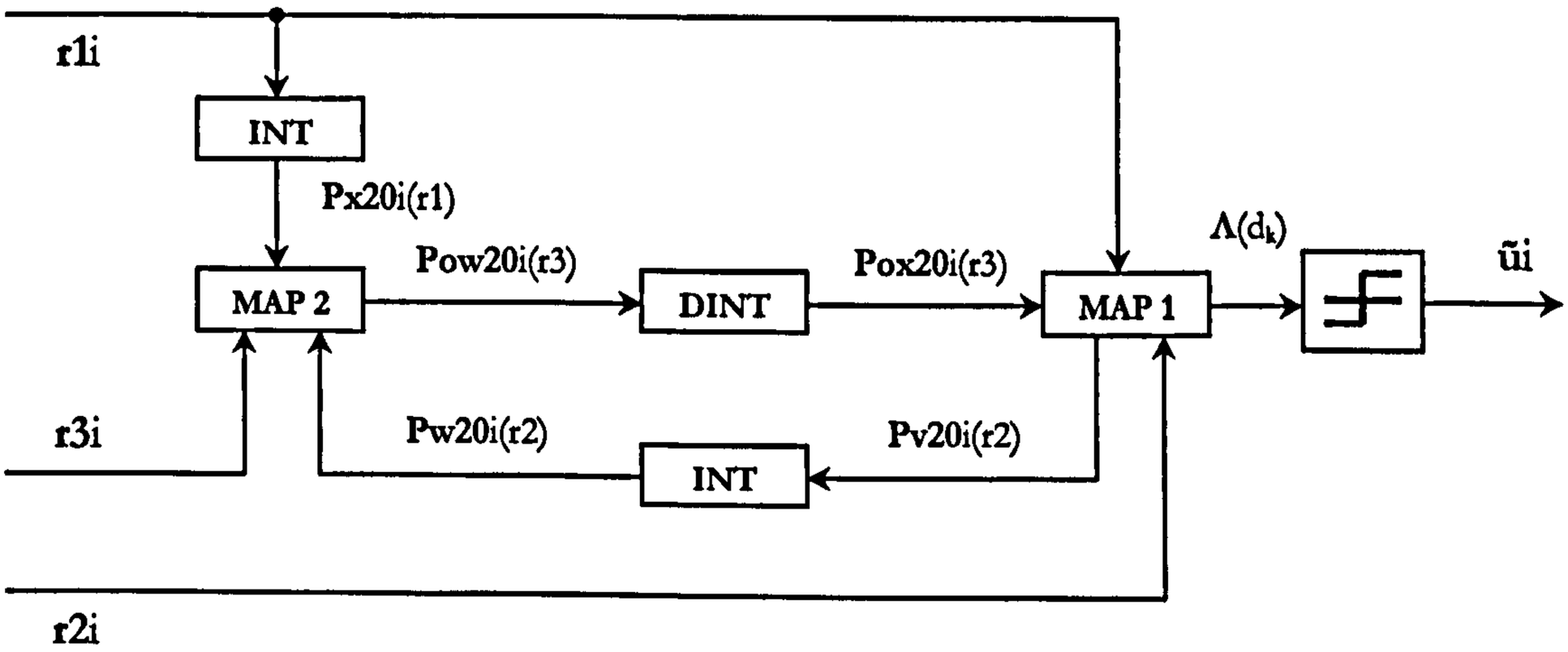


Figure 4-4 The iterative Turbo decoder

begins its operation it already has a better knowledge about the w2i bit since it uses the extrinsic information supplied by MAP2 rather than the start value equal with 0.5. The MAP1 decoder corresponds to the encoder ENC1.

The MAP1 decoder computes the new extrinsic information Pv20i(r2) which is then interleaved and passed back to the MAP2 decoder for a new iteration. The MAP1 decoder also calculates the *a posteriori* probability $\Lambda(d_k)$ by using the extrinsic information Pox20i(r3) and the Gaussian probabilities corresponding to r2i and respectively r3i. When the desired number of iteration was completed, this information is passed to the hardware decision block who gives at its output the error corrected sequence ũi corresponding to the originally encoded data sequence ui.

4.2.2 Several Particularities of the Turbo Codes

It is well known that the performance of the convolutional codes improves with increasing constraint length. This is not the case for Turbo codes. In fact, the best constituent codes of a Turbo code have a very small constraint length.

The performance of convolutional codes does not improve significantly with the decreasing of the code rate; in fact the difference between rate 1/3 and rate 1/128 is of the order of a few tenths of a dB in the case of convolutional codes [Barbulescu et al, 1996]. The Turbo codes instead achieve a very significant coding gain for lower coding rates. For practical code rates between 1/2 and 1/6, **Figure 4-5** illustrates the Eb/No required to achieve a BER of 10^{-6} as a function of coding rate, for both convolutional codes and Turbo codes [Barbulescu et al, 1996]. Analysing the difference in Eb/No between the 1/2 and 1/6 cases for both

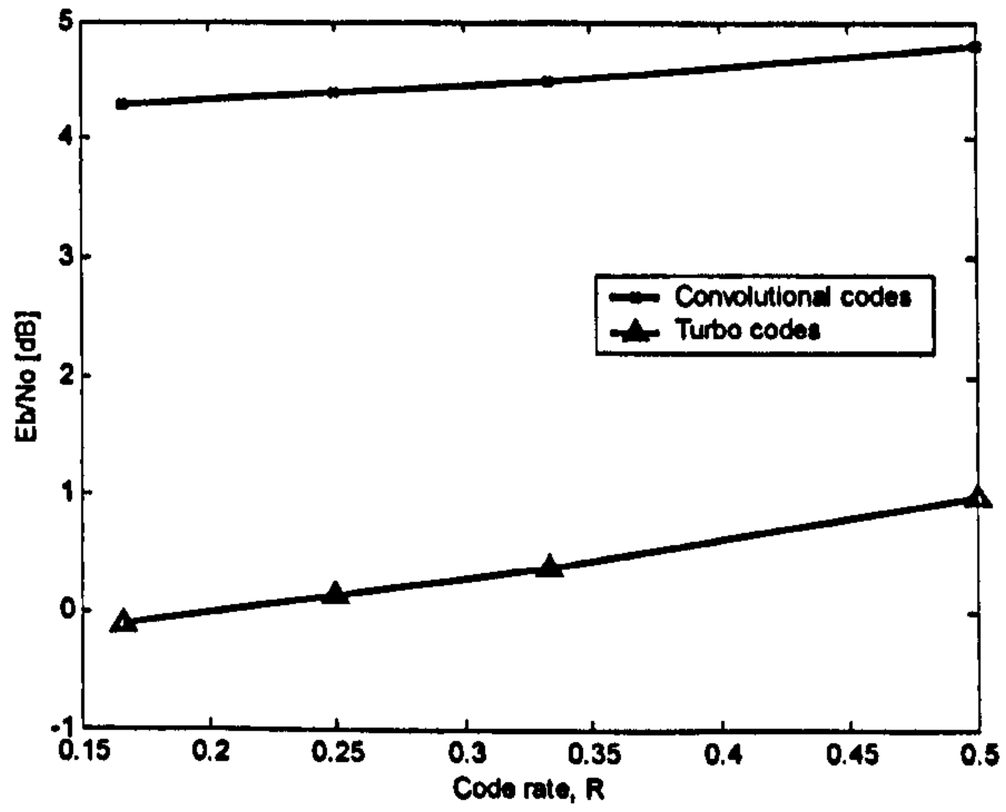


Figure 4-5 Eb/No required to achieve a BER=10^{-6} for convolutional codes and Turbo codes

convolutional codes and Turbo codes it can be seen that this difference is twice as much for the Turbo codes. So it can be concluded that lower rate Turbo codes provide significantly more coding gain (twice as much) than lower rate convolutional codes.

The PCCC are the best choice when the $BER \geq 10^{-6}$, but however for much lower bit rate requirements other codes could be better (SCCC for example). This is because for the PCCC usually a change in the slope of the BER curve appears for $BER < 10^{-7}$ depending of the interleaver size and design. At low Eb/No the PCCC performs better than SCCC but increasing the Eb/No the SCCC schemes outperform PCCC schemes. The crossover point depends again of the interleaver size and design.

Referring strictly to the PCCC it was proven [Divsalar et al, 1995] that the interleaver gain term depends on the number of codes in the concatenated system, and the probability of error is

$$BER \sim \frac{1}{N^{m-1}} \tag{4.19}$$

where N is the interleaver length and m is the number of component codes.

4.3 Turbo Codes in Watermarking

If we regard the watermark channel as a communications system with input X (the watermark data) and output Y, the channel capacity is formally defined as the maximum mutual information (section 4.1.2),

$$C_{chan} = \max_{p(x)} I(X;Y) = \max_{p(x)} \left[H(X) - H(X \mid Y) \right] = \max_{p(x)} \left[H(Y) - H(Y \mid X) \right] \tag{4.20}$$

where the maximum is taken over all possible distributions $p(x)$. Term $H(X \mid Y)$ represents information loss due to channel noise, which will be a combination of the host video and signal processing (compression or other forms of attack). If the loss is modelled as the addition of an independent Gaussian noise source, $Z \sim N(0, \sigma_z^2)$, i.e. $Y_l = X_l + Z_l$, where Z is a continuous random variable, then equation (4.20) reduces to

$$C_{chan} = \frac{1}{2} \log_2 \left(1 + \frac{\sigma_x^2}{\sigma_z^2} \right) \text{[bits/symbol]} \tag{4.21}$$

providing $X \sim N(0, \sigma_x^2)$.

In this thesis the HVS models are employed in order to maximize the signal power. The watermarking channel is illustrated in **Figure 4-6** and can be modelled as a gain factor cascaded with a Gaussian noise source $Z \sim N(0, \sigma_z^2)$ (the gain and variance depending upon the host media, MPEG compression, geometric attack, etc.).

For example, one could estimate a basic operational capacity as follows. Suppose that all N_p pixels ($N_p = 720 \times 576$) in the frame are transformed via the DCT or DWT, and that the channel noise is simply that of the host video. Assuming C_i is i.i.d. for simplicity, the noise power per video frame is $N_p \overline{C^2}$, where $\overline{C^2}$ is the mean coefficient power for a particular video sequence. If only one data bit is embedded per video frame (corresponding to a spread spectrum chip rate $c_r = N_p$), and there is no FEC, the *SNR* is

$$SNR = N_p \frac{\left(\alpha \ \overline{HVS}\right)^2}{\overline{C^2}} = N_p \overline{SNR} \tag{4.22}$$

where $\overline{SNR}$ is a measured mean *SNR* for the video sequence and $\overline{HVS}$ is the mean embedding strength given by the HVS model for the sequence.

If N_b data bits are embedded into a frame (using all the coefficients) the signal to noise ratio per uncoded data bit reduces to $SNR_u = SNR / N_b$, and the data rate or capacity for an uncoded system of frame rate F_r is

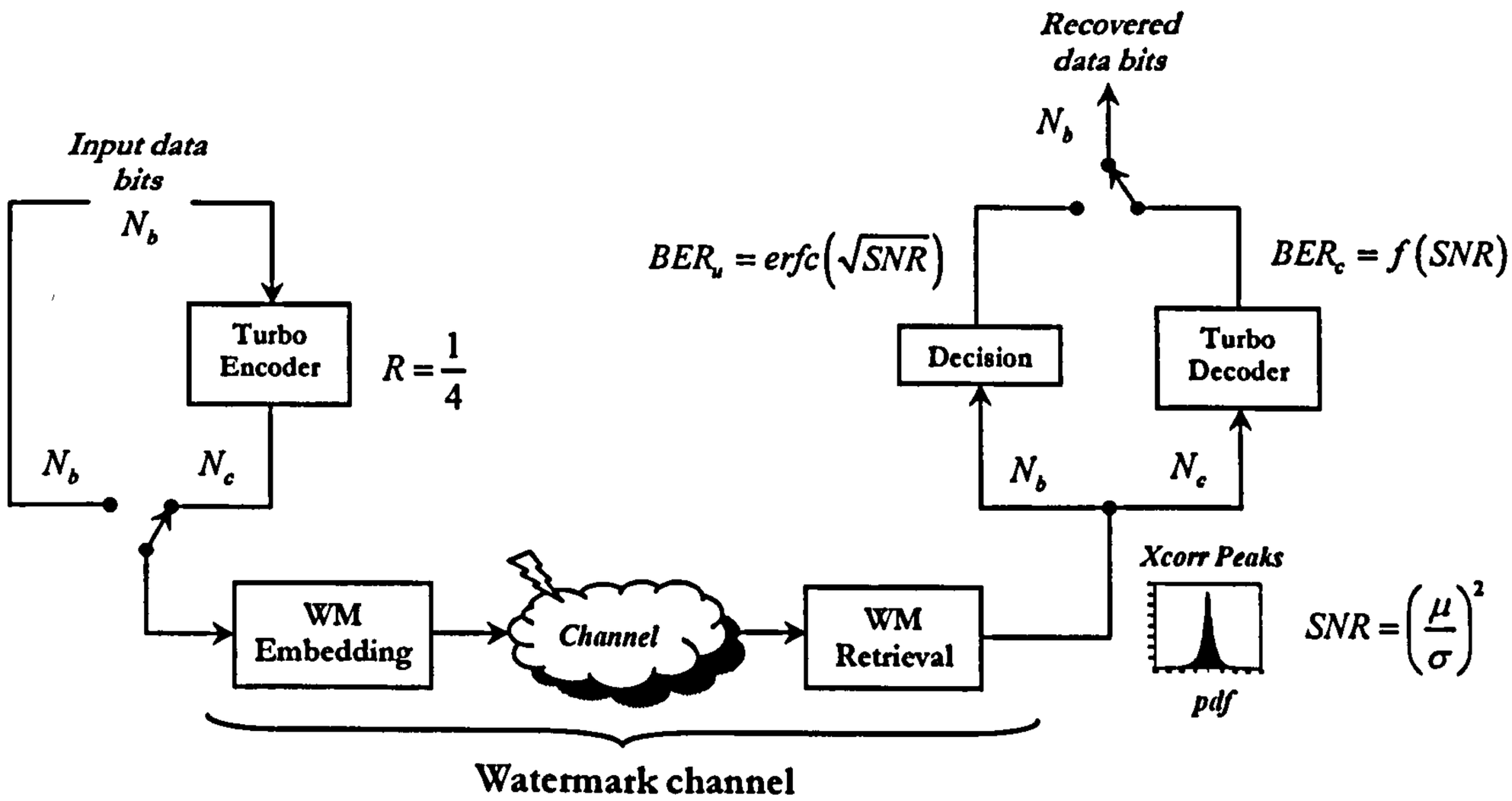


Figure 4-6 The watermarking channel

$$D_r = N_b F_r = \frac{N_p \quad \overline{SNR} \quad F_r}{SNR_u} \tag{4.23}$$

For a coded system of rate R, with $N_c = N_b / R$ bits embedded over N_f video frames, we have

$$D_r = \left(\frac{N_c \cdot R}{N_f} \right) \quad F_r \tag{4.24}$$

The operational capacity can be defined as the maximum value of D_r for which the *BER* does not exceed a tolerable level (typically 10^{-8}).

The *SNR* can be defined as in **Figure 4-6**. Since the cross-correlator performs a sequence of correlation sums, it follows from the Central Limit theorem that the cross-correlation peaks have a normal distribution [Ambroze et al, 2001]. This is very convenient for the iterative Turbo decoder, which generally assumes a Gaussian input. Thus, for any particular system, the distribution mean μ, and variance σ^2 define a *SNR* of the channel

$$SNR = \left(\frac{\mu}{\sigma} \right)^2 \tag{4.25}$$

and the corresponding *BER* for an uncoded system is simply

$$BER_u = Q[\mu / \sigma] = Q\left[\sqrt{SNR_u} \right] \tag{4.26}$$

For a coded system, μ and σ define a signal to noise ratio SNR_c at the decoder input,

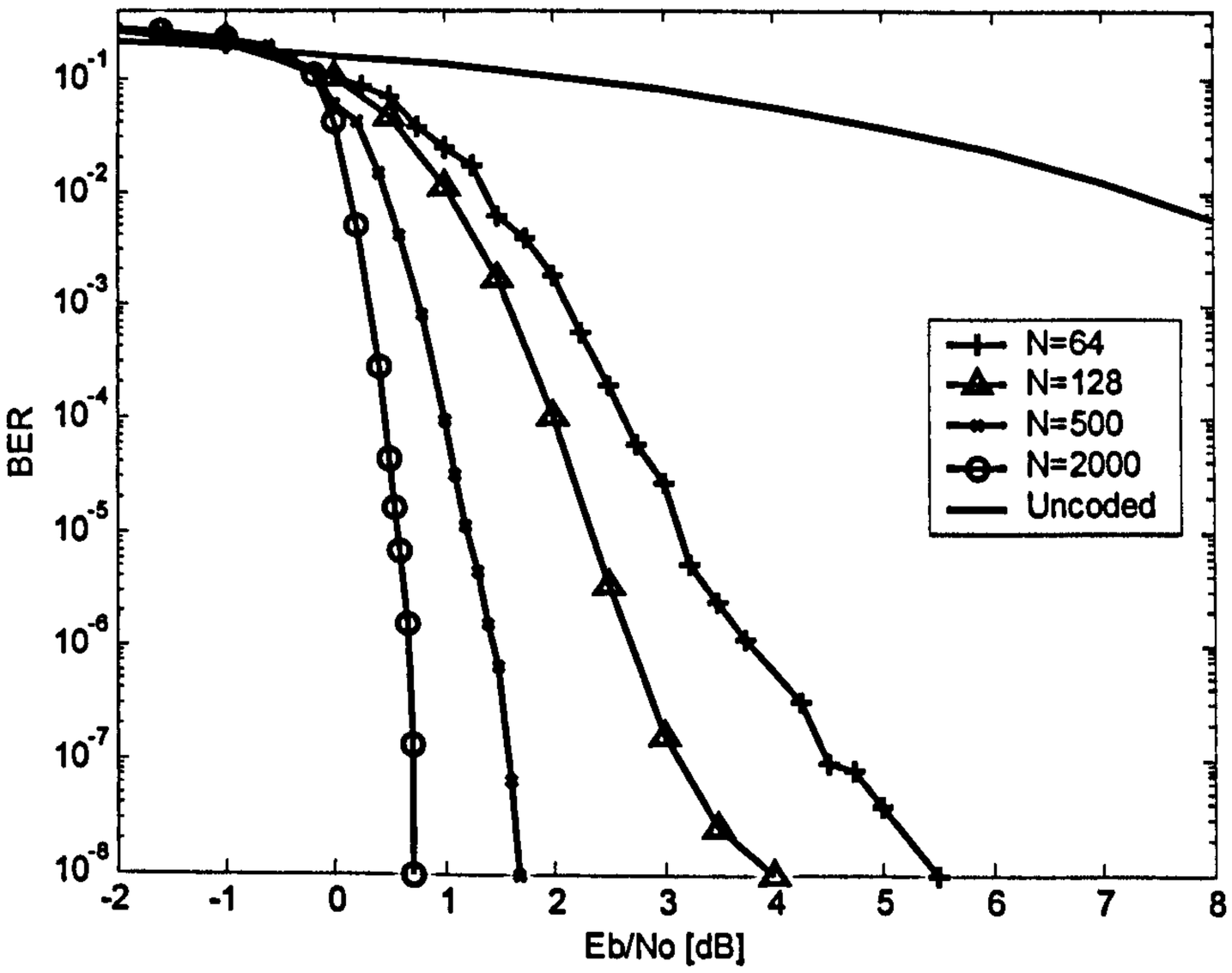


Figure 4-7 The performance of the 3PCCC Turbo code for different block lengths

and the decoded bit error rate is $BER_c = f(SNR_c)$ where f is a known function for a particular iterative decoder.

The FEC code used in this thesis is a rate 1/4 multiple parallel concatenated convolutional code (3PCCC) [Ambroze, 2000] rather than the basic turbo code (2PCCC) in order to improve performance [Ambroze et al, 2001]. The performance of this code is presented in **Figure 4-7**. The structure of the 3PCCC Turbo encoder and decoder is presented in **Appendix 1**.

The use of FEC reduces the chip rate by a factor R due to the fact that now we have to embed $N_c = N_b/R$ coded bits instead of N_b. This increases the variance of the channel distribution, resulting in increased BER, and the FEC decoder must more than compensate for this increase in order to provide coding gain. As it will be shown in the next chapters, the Turbo code improves the performance of the watermarking system in a significant manner.

4.4 Conclusions

The watermarking channel is a very difficult channel characterised by high levels of noise (the video sequence itself represents the noise) and low power of the watermark signal (due to the visibility constraints). The situation is even worse when taking into account various attacks which increase even further the noise from the system. This translates to a relatively low SNR at the input of the FEC decoder.

This fully justifies the use of soft decision based FEC codes, particularly the use of Turbo codes which are known for their very good performance under difficult conditions (very low SNR). Using other error correction codes like the BCH codes (used by some authors in their watermarking systems) which have a hard decision algorithm, automatically leads to a 2dB drop in performance when compared with Turbo codes, or other soft decision based algorithms. This is a very significant loss in a watermarking system, which cannot be afforded.

Digital watermarking seems to be yet another successful application area for Turbo codes, joining many other already consecrated application areas like deep space applications, satellite communications, speech and image transmission and many others.

Chapter 5

"I have the terrible feeling that, because I am wearing a white beard and am sitting in the back of the theatre, you expect me to tell you the truth about something. These are the cheap seats, not Mount Sinai."

George Orson Welles (1915-1985)

Watermarking in the DCT Domain

The literature largely agrees that watermarking in the transform domain offers higher capacity and increased robustness compared to the spatial domain. This chapter presents the case of watermarking in the DCT domain, together with several methods of increasing the capacity/robustness of the system. To achieve this goal, the system uses both advanced HVS models for watermark embedding and state-of-the-art FEC (Turbo codes) in order to protect the watermark. The casting of the watermark and other alternative modulation techniques are also analysed. A description of the DCT transform and its properties, together with many other references can be found in [Jain, 1989].

In order to improve the system even further, 3-D marking replaces the usual frame by frame approach (2-D marking) by taking into account the temporal dimension. This increases the "local" chip rate leading to better cross-correlation results (wider cross-correlation area) and caters for frame dropping/duplication attacks.

5.1 Watermark Embedding in the DCT Domain

The DCT based watermark embedding is presented in **Figure 5-1**. The scheme is similar to the spatial domain approach, but it has several differences due to the particularities of the DCT domain marking.

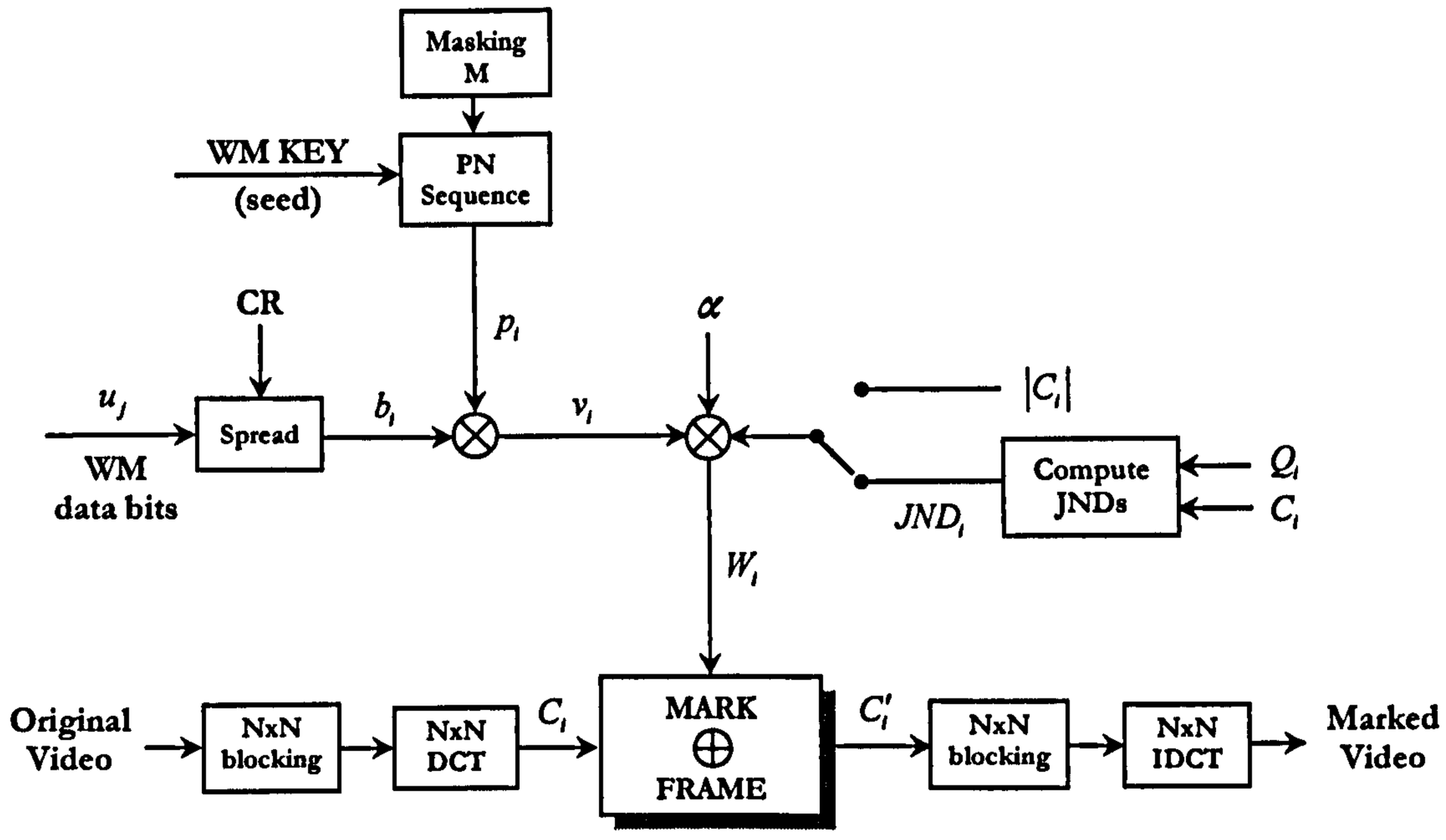


Figure 5-1 Watermark embedding in the DCT domain

The strength of marking is given for each DCT coefficient by an advanced visual model represented in **Figure 5-1** as "Compute JNDs". This block calculates the so called *Just Noticeable Difference* (JND) measure which represents the maximum value which can be added or subtracted from the given DCT coefficient, without leading to perceptual artefacts in the marked sequence. In other words, the HVS model keeps the strength of the marking just below the visibility threshold and ensures that the sequence is marked with the maximum energy and yet the invisibility requirement is still satisfied. The factor α is used as a global adjusting factor for the entire frame, either to attenuate or to amplify the value given by the HVS model for some difficult sequences.

There is also the possibility of using a "heuristic" marking as well, where the amplitude of the watermark is directly proportional to the amplitude of the DCT coefficient. The results in this case are much worse.

The watermark generation takes into account the masking matrix M, which allows one to select which coefficients within the DCT block will be marked and which skipped. The matrix M has the same dimension to the DCT block and can only have two values: zero and one. Value zero signifies that the respective coefficient is skipped (not marked) and value one that the corresponding coefficient is marked. Some authors suggests that only the medium frequency DCT coefficients should be marked since the low frequency coefficients lead to visibility artefacts and the high frequency coefficients are not robust to compression attacks. The experiments show that in fact is actually better to watermark all DCT coefficients rather

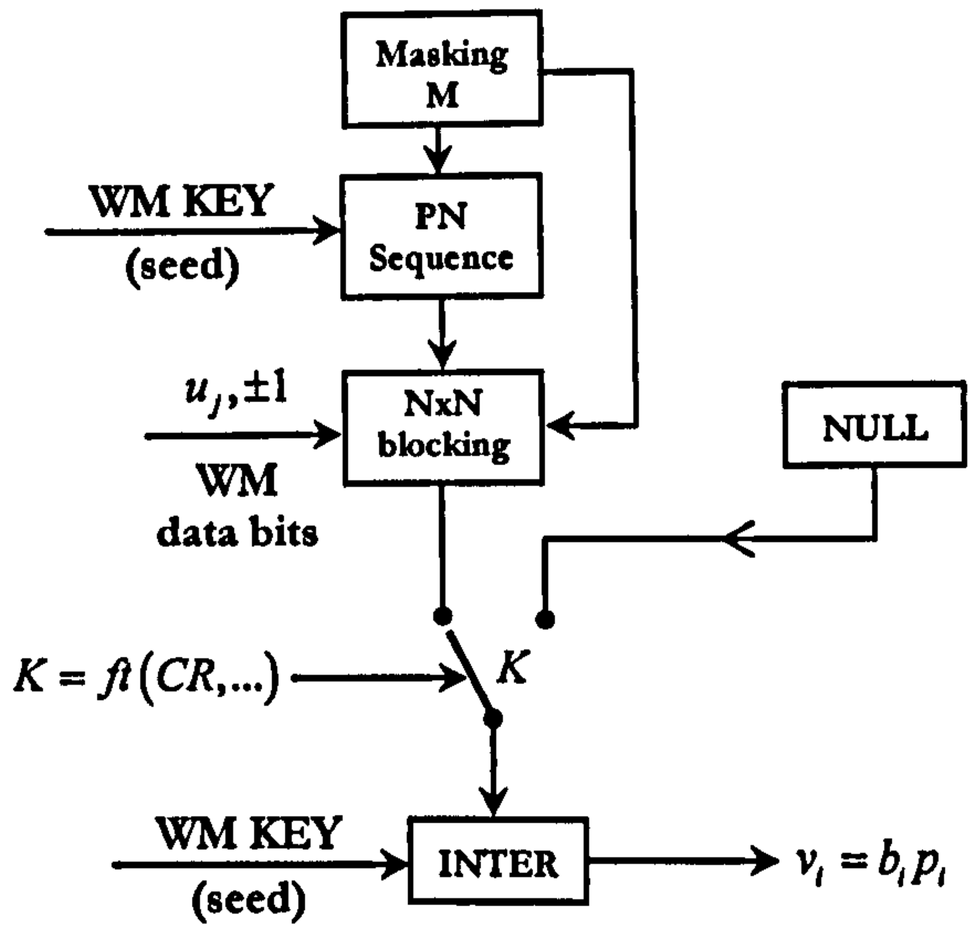


Figure 5-2 Watermark spreading detail

than excluding some of them. So, although this feature could be useful for the "heuristic" schemes, the use of advanced perceptual models (as the JND is) makes it rather inutile. In fact when is used together with the JND model, the performance of the system decreases. The HVS model inserts so much energy in the high frequency coefficients that those coefficients are quite robust to compression attacks and is a waste not to mark them. This is the reason for setting all the values within the matrix M to one.

More details about the casting of the watermark are illustrated in **Figure 5-2**. Several things can be observed. There is the possibility of watermarking only a desired number of blocks. In this case, depending of the desired percentage of marked blocks (which gives the actual chip rate) some blocks can be skipped (e.g. "marked" as null blocks).

In order to save time and memory resources, the PN sequence is generated only for the valid positions within the block (where the value of matrix M is one). Also, the PN sequence is generated only for the valid blocks (not for the null blocks). The PN sequence is generated using the same multiplicative congruential generator described in section 3.1.1.

Since the security of the algorithm is very important, the PN sequence is generated according with a secret key. The security of the algorithm is further improved by using an interleaver block INTER, also dependent on a secret key (it can be a different key or the same key used for generating the PN sequence). The use of the interleaver ensures a pseudo random distribution of the watermark data bits within the frame and within the consecutive frames as well. Each NxN DCT block corresponds to one input watermark data bit.

The watermark is embedded according to the following equation

$$C'_{n,i} = C_{n,i} + w_{n,i} = C_{n,i} + \alpha b_{n,i} p_{n,i} JND_{n,i} \tag{5.1}$$

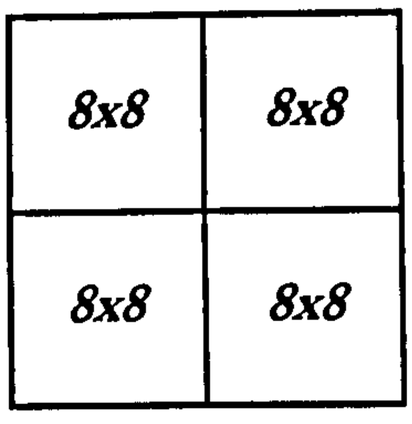


Figure 5-3 Structure of the macro-block

where $C'_{n,i}$ represents the $i'th$ coefficient from the marked block n, $C_{n,i}$ represents the original coefficient, α is an amplitude adjusting factor, $b_{n,i}$ represents the spread input data bit corresponding to block n, $p_{n,i}$ is the pseudo-random sequence corresponding to this block and finally $JND_{n,i}$ contains the HVS values associated with the block n.

5.1.1 Block Sizes and Macro-Blocks

It is well known that the cross-correlation process gives better results for larger cross-correlation areas. In the case described above, the cross-correlation area is relatively small: NxN. Typical values for N are 8, 16, 32, 64 and 128. Unfortunately N cannot be too large because of the desired resilience to attacks like line or column cuts. On the other hand the JND model works well only for small blocks, giving the best results for 8x8 blocks. To overcome at least partially this problem, one can introduce the concept of *macro-block*.

As **Figure 5-3** shows a macro-block is composed of four additional 8x8 DCT blocks coupled together. Using 8x8 blocks ensures that the HVS model works at its full potential, and by connecting four of these blocks together the effective chip rate increases four times and therefore the cross-correlator works better (the cross-correlation window is in this case 16x16). In other words, the marking is done on 8x8 blocks and the recovery of the watermark on 16x16 blocks. A macro-block corresponds to one input data bit.

5.1.2 PN Sequence Arrangement

The idea is to use the same PN value for a group G of additional DCT coefficients, in the hope that this kind of arrangement will be more robust to geometric attacks and it will improve the overall performance of the system. For the typical case described in the thesis (8x8 blocks) four additional DCT coefficients were used (in a square shape). For each of these four coefficients has been assigned the same PN sequence.

The experimental results show that by using such an arrangement, the performance of the system is slightly better (the SNR of the peaks is slightly bigger). In fact it has been observed that the variance of the peaks tends to decrease but the mean of the peaks decreases as well, although not as quickly as the variance. Generally speaking the difference between this case and normal marking is only about 1%. Although this difference is small, since this technique does not increase the complexity of the algorithm, it still constitutes a gain.

5.1.3 Alternative Modulation Techniques

The modulation used until now can be described as "amplitude" modulation since the process involves a simple addition of the watermark. On the other hand, it is well known from communication theory that differential modulation is superior to amplitude modulation. Based on this assumption one could try to implement a "differential" modulation technique for watermarking and profit from its superior noise immunity. Such an idea called "cocktail watermarking" was described in [Lu et al, 1999 and 2000] for a non-blind system. As usual, for blind systems the problem is always more difficult but the technique could be adapted to the requirements of a blind watermarking system.

Taking advantage of the macro-block structure already defined in section 5.1.1, one could adapt it for differential modulation. As **Figure 5-3** shows, there are 4 blocks available within a macro-block, and one macro-block has associated one watermark data bit.

One idea could be to divide the macro-block into two or four regions as **Figure 5-4** suggests. In the first two cases the macro-block is divided in two areas. Lets assume that the watermark data bit corresponding to this macro-block is u_j. Then the first half of the macro-block can be marked with u_j and the second half with $-u_j$ (or the other way around). At the retrieval two distinct correlations are required for each macro-block: one for the first half of the macro-block and the other one for the second half. The decision is taken by comparing the

$+u_j(-u_j)$	$-u_j(+u_j)$
$+u_j(-u_j)$	$-u_j(+u_j)$

$+u_j(-u_j)$	$+u_j(-u_j)$
$-u_j(+u_j)$	$-u_j(+u_j)$

$+u_j(-u_j)$	$-u_j(+u_j)$
$-u_j(+u_j)$	$+u_j(-u_j)$

Figure 5-4 Differential modulation

signs of those two regions. The amplitude of the peak can be used as well, as a confidence measure. Assuming no errors, the signs of those two regions are always different. If the signs are not different or if the peaks are not high enough one could decide to discard the cross-correlation result entirely as unreliable.

The third case illustrated in **Figure 5-4** can be interpreted either as in the first two cases, considering in this instance a diagonal splitting of the block, or one could consider the block divided into four regions. Similar principles can be applied in this case too; the only difference is that now, four distinct correlations are required for each macro-block.

Using this kind of differential marking has however an important drawback: the effective chip rate decreases two (or respectively 4) times; in other words, the cross-correlation area is two or four times smaller, which has a negative effect on the cross-correlator.

The experiments carried out for the various arrangements described in **Figure 5-4** show that overall the scheme performs marginally worse than the normal scheme (with amplitude modulation). This is due to the smaller effective cross-correlation area. As a conclusion, taking into account that the scheme is slightly more complex and performs slightly worse, one should stick with the normal "amplitude" modulated scheme.

5.2 The Just Noticeable Difference

A very successful application for perceptual models has been proven to be image/video compression [Jayant, 1993-1 and 1993-2]. Perceptual models allow one to take advantage of the characteristics of the HVS in order to remove irrelevant and redundant information whilst keeping the compression artefacts as low as possible.

One of the most advanced HVS models, the JND model, was developed by Watson [Peterson et al, 1993], [Ahumada et al, 1992], [Watson, 1993], [Watson et al, 1994]. The aim of this model is to provide a (down to the coefficient) adaptive quantisation matrix for a JPEG based encoder. The JND algorithm is superior to many other HVS models already mentioned in section 2.4.2, due to its highly adaptive nature. This HVS model supplies a (different) JND value for each DCT coefficient.

The perceptual model used in this chapter is based on a simplified form of this HVS model. Using the idea presented in [Kim et al, 1999] this algorithm is extended to account for yet another masking effect of the HVS. The algorithm is described below.

5.2.1 Modulation Transfer Function

The model starts by computing first the *frequency sensitivity* of the eye as described by the *modulation transfer function* of the eye (MTF). The MTF describes the human eye's sensitivity to sine wave gratings at various frequencies and provides only a basic approximation of the visual model, that depends only on the viewing distance, equipment and other viewing parameters and it is independent of the image content. The result can be interpreted as a static JND threshold for each frequency band. An example could be the classical quantisation matrix from the JPEG standard.

The model developed by Watson computes this threshold using a complex formula which involves several parameters dependent on the viewing conditions. Since for watermarking the goal is only to compute the JND threshold rather than the perceptual quantisation matrix as in Watson's case, it is possible to simplify this step significantly

$$T_F(i) = \frac{Q_i}{2} \tag{5.2}$$

where Q_i is the standard quantization matrix of the JPEG standard (or any other quantization matrix developed for JPEG). This simplification affects only marginally the performance of the algorithm.

5.2.2 Luminance Masking

The next step is to compute the *luminance masking (sensitivity)* threshold. Luminance sensitivity is a way to measure the effect of the detectability threshold of noise on a constant background. This phenomenon depends on the average luminance value of the background as well as the luminance level of the noise. It basically suggests that the noise is more visible on a low intensity constant background than a high intensity contrast background. For the HVS system this is a nonlinear function. Since luminance sensitivity takes advantage of the local luminance levels from the image/video it is important that the size of the block is small enough. The luminance sensitivity is defined in [Watson, 1993] as follows

$$T_L(i,k) = T_F(i) \cdot \left[\frac{C_{0,k}}{\overline{C}_0} \right]^{0.649} \tag{5.3}$$

where $C_{0,k}$ represents the DC coefficient within block k and $\overline{C}_0$ corresponds to the mean DC coefficient over a frame.

5.2.3 Contrast Masking

A further refinement can be achieved by extending the visual model to include *contrast masking*. Contrast masking refers to the detectability of one signal in the presence of another signal (noise, artefacts). The effect is strongest when both signals are of the same spatial frequency, orientation and location [Legge et al, 1980]. More complex regions can tolerate more distortion than a smooth region or a region containing a simple sharp edge. The contrast masking is computed as in [Watson, 1993]

$$T_C(i,k) = T_L(i,k) \cdot \max\left[1, \left(\frac{|C_{i,k}|}{T_L(i,k)}\right)^w\right] \tag{5.4}$$

where $C_{i,k}$ represents the $i'th$ DCT coefficient from block k and $w = 0$ for the DC coefficient and 0.7 elsewhere.

Equation (5.4) accounts for three important components of the human visual system: frequency, luminance and respectively contrast sensitivity. It can be seen that each new sensitivity threshold depends on the previous one.

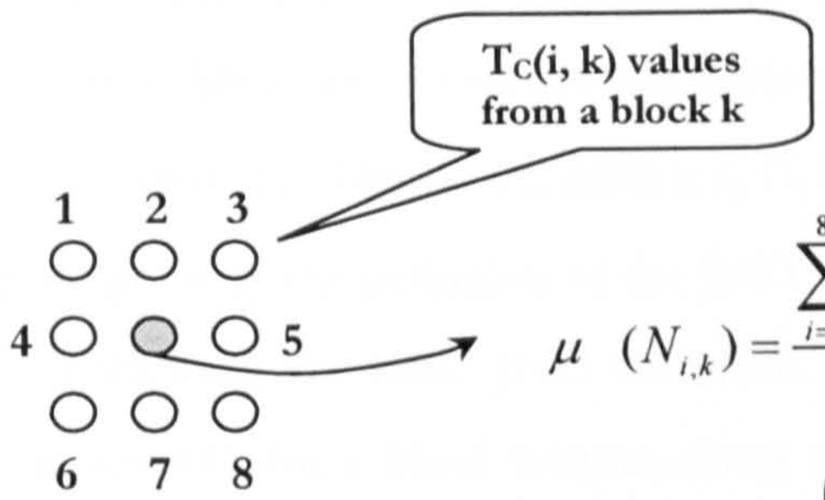
5.2.4 Lateral Inhibition Masking

Using the same idea as in [Kim et al, 1999], the model can be extended by incorporating another masking effect, the *lateral inhibition masking*

$$T_{LI}(i,k) = \left\{ \begin{array}{l} T_C(i,k) \text{ if } \left(T_C(i,k) > \mu(N_{i,k}) \text{ or } (T_C(i,k) - \sigma(N_{i,k})) < \frac{Q_i}{32}\right) \\ T_C(i,k) - \sigma(N_{i,k}) \text{ otherwise} \end{array} \right\} \tag{5.5}$$

where $\sigma(N_{i,k})$ and $\mu(N_{i,k})$ are the standard deviation and respectively the mean for the eight neighbours of $T_C(i,k)$ and can be calculated as in **Figure 5-5**.

In the HVS, the horizontal and amacrine cells transmit signals to the neighbouring bipolar and ganglion cells, which inhibit their responses. Lateral inhibition model simulates this characteristic of HVS. In this way, the use of equation (5.5) ensures that greater marking energy will be assigned to those coefficients that are susceptible to the inhibitory effect of their neighbours. The condition from equation (5.5) was obtained from limited subjective visibility tests carried out under this project.



$$\mu\ (N_{i,k}) = \frac{\sum_{i=1}^{8} T_C(i,k)}{8}$$

Mean of 8 neighbours of the current threshold

$$\sigma\ (N_{i,k}) = \sqrt{\frac{\sum_{i=1}^{8} \left(T_C(i,k) - \mu(N_{i,k})\right)^2}{8}}$$

Standard deviation of 8 neighbours of the current threshold

Figure 5-5 Computing the parameters of lateral inhibition masking

5.2.5 JND Threshold

Other authors [Podilchuk et al, 1997-1, 1997-2 and 1998], [Wolfgang et al, 1999], [Kim et al, 1999] use $T_C(i,k)$ or a form of $T_{LI}(i,k)$ directly as JND values, and marking is *conditional* i.e. a condition based on the original frame dictates which coefficient can be marked and which

Figure 5-6 The JND map (profile) of the Lena image

cannot. In these schemes the original frame can be used to determine the marked coefficients (they are non-blind) and so watermark retrieval is possible, but this approach is not feasible for video watermarking. Moreover, neither $T_C(i,k)$ or $T_{LI}(i,k)$ are strictly speaking JND values in the sense given by the definition of the JND.

Following the basic JND definition given in Watson's paper, leads to *unconditional* marking suitable for a blind watermarking system. Moreover this approach is less empirical than the previous method and gives better results. The JND values can be computed as

$$JND_{i,k} = \frac{Q_i}{2T_{LI}(i,k)} \tag{5.6}$$

Clearly, JND values are both HVS and media dependent. In practice, the theoretical JND values supplied by equation (5.6) are within a factor 2 or 3 below of the actual perceptual threshold, and this is accounted for by the factor α in equation (5.1).

The JND map of the well known image Lena is presented in **Figure 5-6**.

5.2.6 Advantages of the JND Model

Using a good HVS model constitutes a requirement for any watermarking system; the use of such a model significantly improves the robustness and the capacity of a watermarking system. This is particularly true for highly adaptive HVS models as JND model is, and can lead to optimal embedding strength. Summarising, one can say that the JND model:

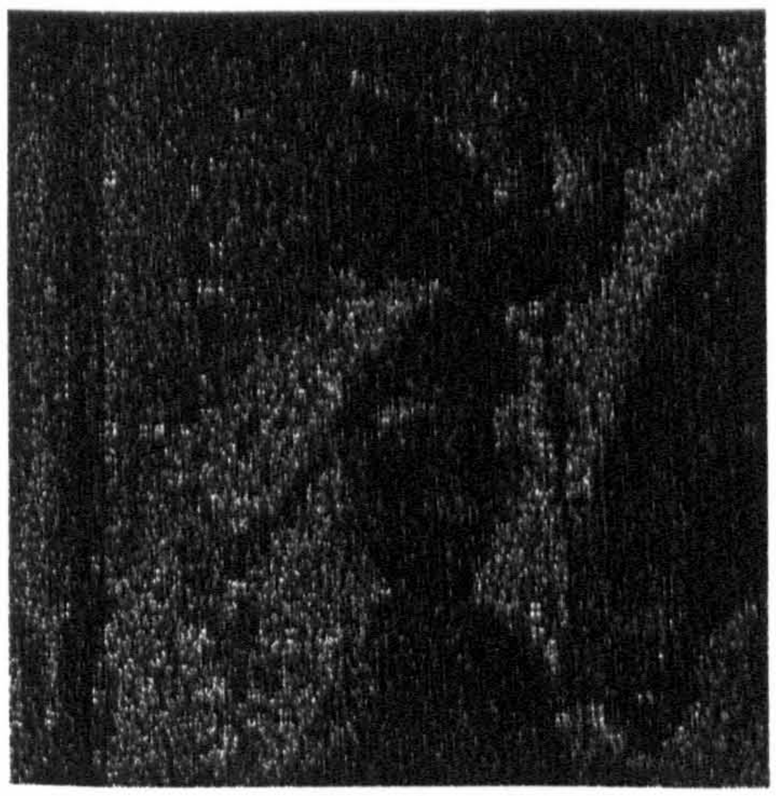
- Exploits various properties of the HVS and adaptively controls the amount of watermark energy to be embedded into each transform coefficient of the image/video; in other words the algorithm is both *HVS dependent and media dependent.*
- *Provides an upper bound* on the amount of modification one can make to the content of original image/video without incurring perceptual differences. The algorithm proves to be reasonably *accurate* for a variety of images/video sequences.
- Provides the maximal strength of the watermark which can be embedded into an image/video, leading to *maximal robustness, capacity and invisibility.*

The disadvantage of the JND model is its relative complexity, but its use can be justified by the good performance of the algorithm.

(a)

(b)

(c)



(d)

(e)

Figure 5-7 Lena image: **(a)** the original, **(b)** JND based watermarked version, **(c)** "heuristically" watermarked version, **(d)** the watermark corresponding to image (b) and **(e)** the watermark corresponding to image (c).

5.2.7 Examples of Watermarked Images

To illustrate the power of the visual model **Figure 5-7** shows the exaggeratedly marked image Lena and the corresponding watermark for both JND based marking and "heuristic" marking for similar performance results (BER). For the heuristic case, the factor α was chosen to be $\alpha = 0.06$ and for the JND marking case this factor was set to $\alpha = 6$. Both images representing the watermark were amplified 8 times in order to see the watermark properly.

One can see that the level of distortion is much higher and the artefacts much more annoying for the "heuristic" marking case. The JND marking instead leads to a noise like type of the visual artefacts which is much easily tolerated by the human visual system. By analysing the watermark itself, defined as the difference between the marked image and the original image, it can be easily seen that the "heuristic" marking leads to a much coarser watermark which does not account for the specifics of the HVS.

5.3 Watermark Recovery in the DCT Domain

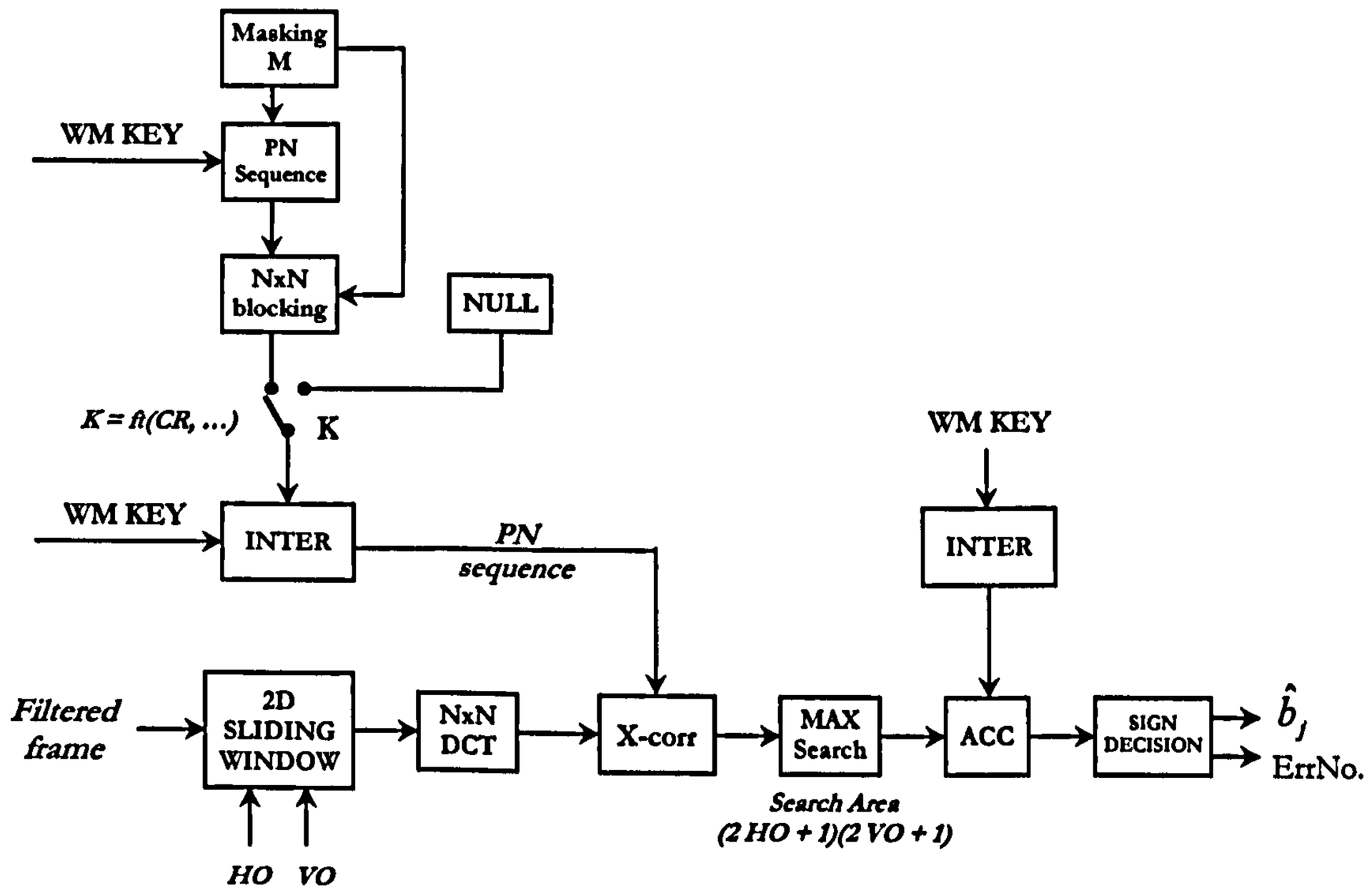


Figure 5-8 DCT watermark retrieving

The recovery of the watermark is very similar to the spatial domain technique described in section 3.2.2. The schematic of the recovery process is presented in **Figure 5-8**.

The frame is first filtered using a Laplacian filter. Then the 2-D sliding window block reads the appropriate block (macro-block) from the frame and the DCT transform of this block (macro-block) is performed obtaining the DCT coefficients.

The cross-correlation between these coefficients and the same PN sequence used for embedding (given by the watermarking key) is computed and compared to the other partial cross-correlation peaks obtained for all the other possible sliding positions.

When all these partial sliding results corresponding to one block (macro-block) are computed the maximum value is delivered to the accumulator ACC which adds this value to the corresponding previous value for that particular input bit. It can be noticed that the accumulator buffer has the same length as the number of input watermark data bits.

After all the blocks within a frame and all the frames were processed in this way, the accumulator will contain the final cross-correlation peaks for al the input data bits. Finding the (estimated) value of the input bit involves a simple sign decision (with the threshold set to zero).

The role of the second interleaver is to communicate to the accumulator the correct position of the current bit in the ACC buffer. In fact the second interleaver together with the accumulator acts as a "deinterleaver".

5.4 Temporal Dimension: 3-D Sliding Correlator

5.4.1 Temporal Macro-Blocks

The advantage of having a bigger cross-correlation area was already discussed in section 5.1. Until now, the solution was to use the macro-block as the smallest unit corresponding to (containing) one watermark data bit. Such a spatial macro-block composed by 4 additional 8x8 DCT blocks increases the "local" effective chip rate by a factor of four and therefore the cross-correlator works better.

This approach can be further extended to the temporal dimension. **Figure 5-9** illustrates this concept. By using time as the third dimension, one could extend the concept of macro-block to account for this dimension as well.

Therefore a temporal macro-block is composed from a "marking depth" number of spatial macro-blocks as defined in **Figure 5-9**. The "marking depth" can be chosen as a compromise between the size of the cross-correlation window and resilience to time synchronisation attacks. Experiments show that a marking depth of four frames is a good compromise. In this case, when 2-D sliding is performed, the effective block size is 16 times bigger compared to the case of a single block, or 4 times bigger compared to the case of a spatial macro-block which translates to increased performance for the cross-correlator.

5.4.2 Temporal Sliding Correlator

One disadvantage of the scheme described until now is the lack of robustness to time synchronisation attacks like frame dropping or frame duplication. This flaw of the existing scheme can be addressed by using a 3-D cross-correlator rather than the 2-D correlator described before. In this case the search is carried out in 3 dimensions: both in space (2-D) and in time.

The temporal sliding is illustrated in **Figure 5-10**. In this case becomes possible to perform the temporal sliding by moving all the blocks within the frame which correspond to the same watermark data bit in the same time. In this way the effective size of the cross-correlation window is much bigger compared to the case of spatial (2-D) sliding, where this technique cannot be successfully applied because of the "discrete" geometric attacks like line and column cuts. Because of this technique the frame dropping attack can be recovered with minimal loss and therefore the results of temporal sliding are much better compared with those obtained for spatial sliding.

One disadvantage of the technique is complexity. The problem was hard enough for

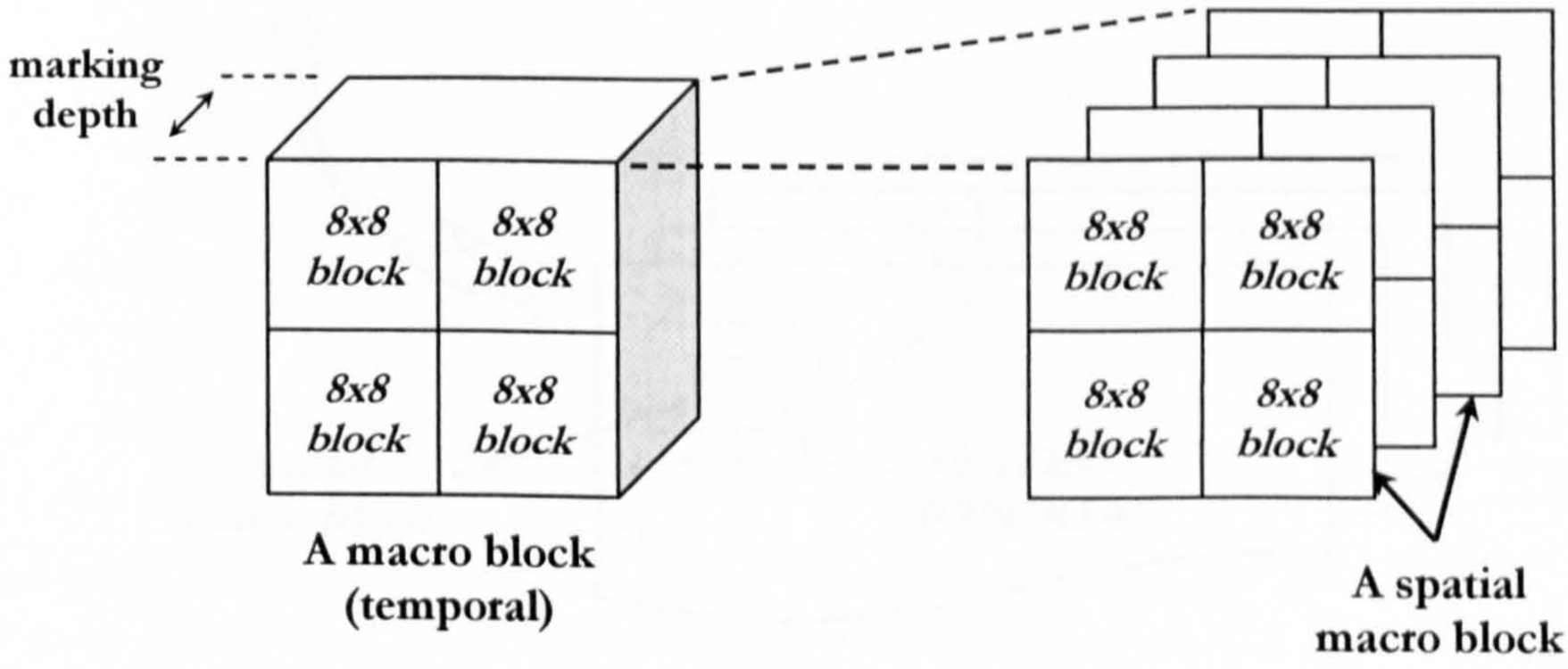


Figure 5-9 Structure of temporal macro-block

the 2-D case, as equation (3.11) shows, but now is even more complex, once the third dimension is added to the equation

$$NC = (2 \cdot ho + 1)(2 \cdot vo + 1)(2 \cdot to + 1) \tag{5.7}$$

where NC represents the number of cross-correlations and ho, vo and to are the horizontal, vertical and respectively the temporal offsets. If for a 2x2 spatial sliding, the 2-D cross-correlator has to perform $NC = 25$ cross-correlations for each block, in this case, assuming a 2x2x2 sliding, the 3-D correlator has to perform $NC = 125$ cross-correlations.

5.5 Performance of the DCT Scheme

The performance of the 2-D DCT scheme with and without sliding is illustrated in **Figure 5-11**. **Figure 5-11(a)** shows the performance of the system for different test sequences,

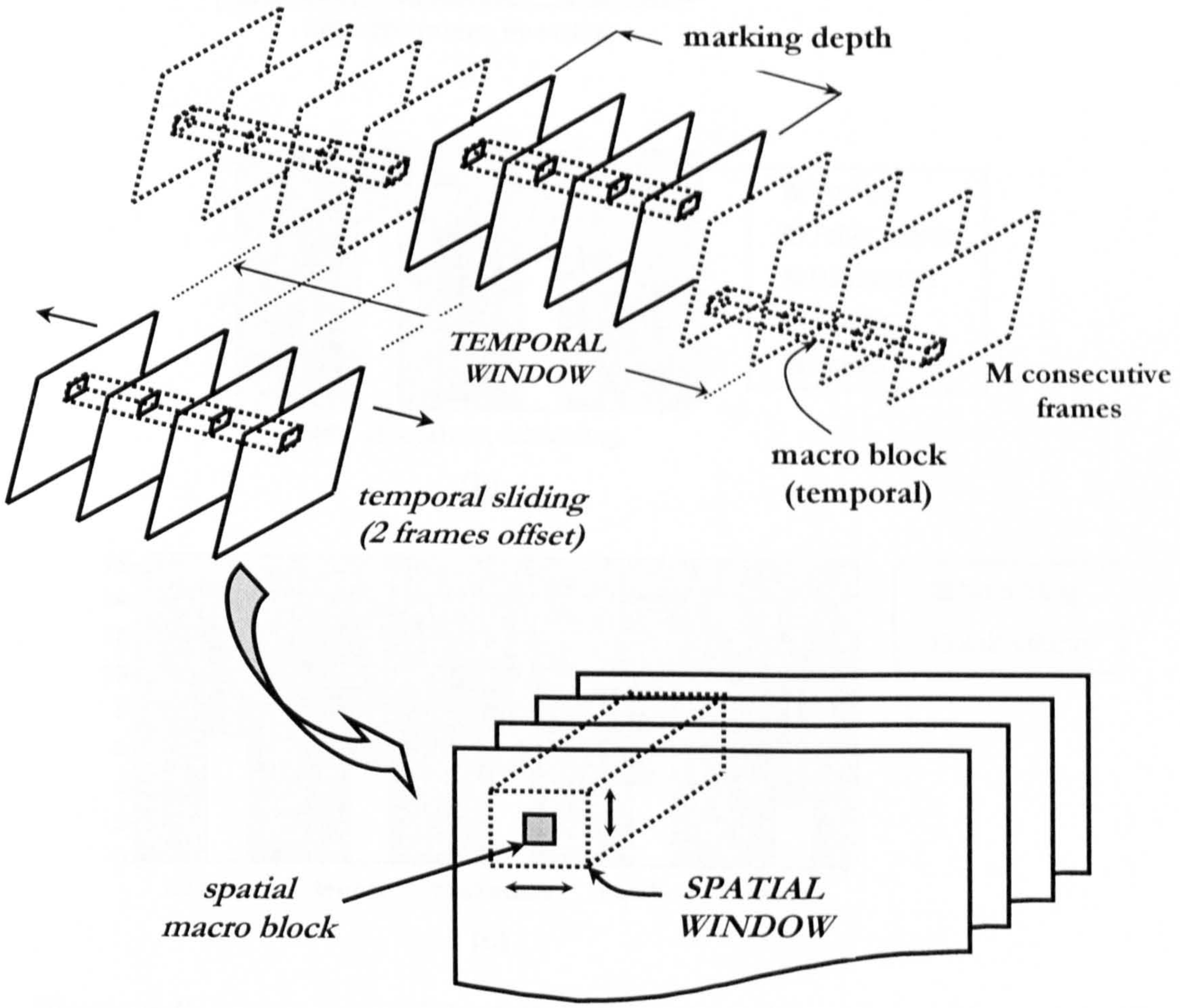
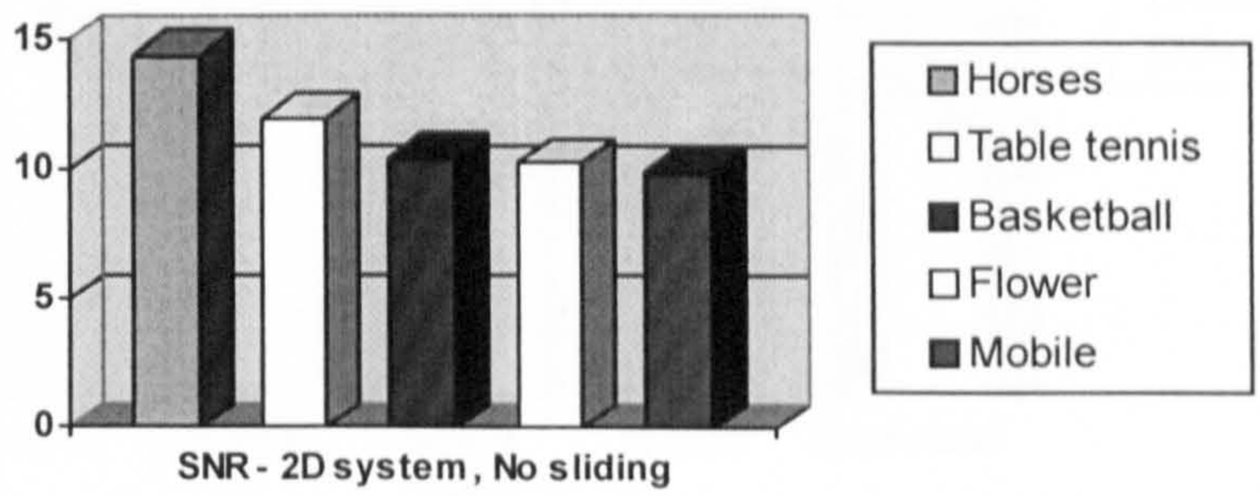


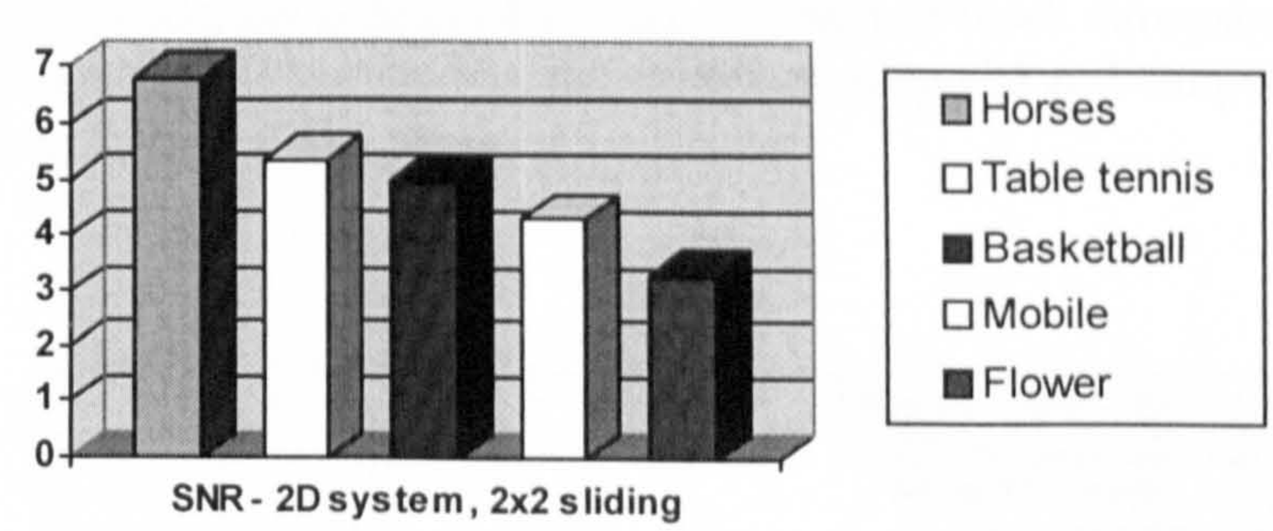
Figure 5-10 Temporal sliding

without sliding. The results for the same sequences but with 2x2 sliding are presented in **Figure 5-11(b)**. The difference between these two cases is shown in **Figure 5-11(c)**.

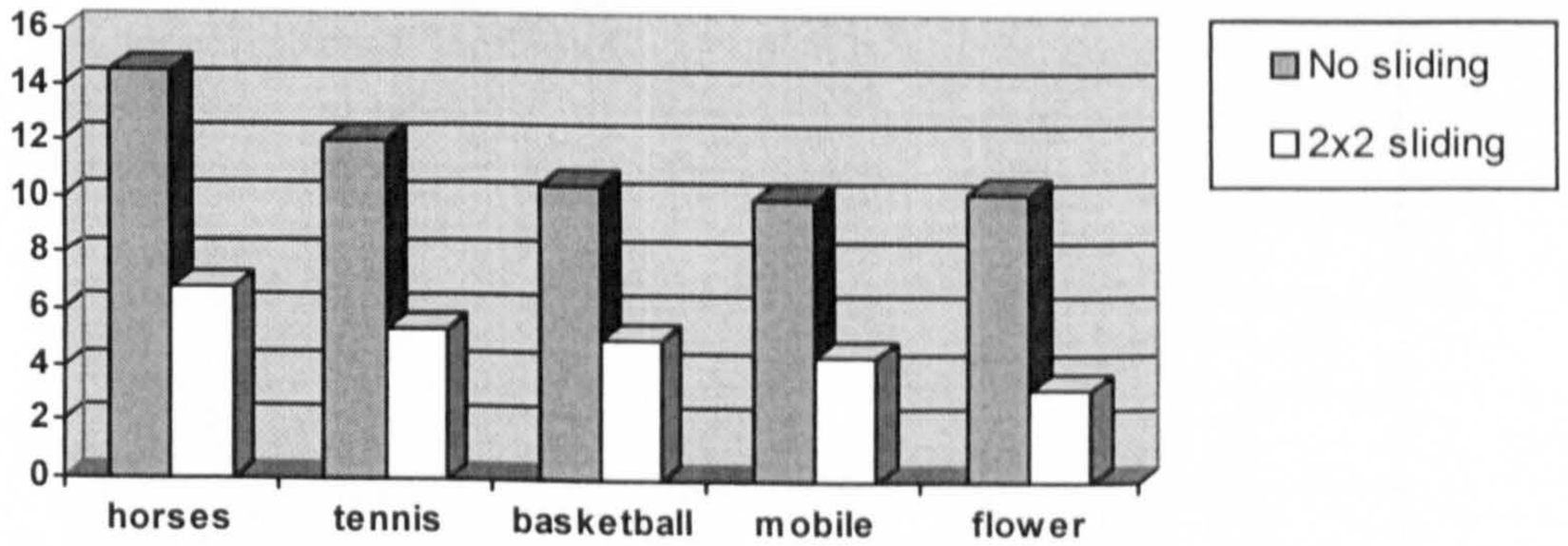
These results were obtained for watermarking 24 video frames with 1024 watermark data bits, for a factor $\alpha = 2$ and without any DCT coefficient masking. The chip rate in this case was 9720, corresponding to 100% block marking percentage (all the blocks were marked). The same parameters were used for **Figure 5-12**. For the 3-D scheme the marking depth was set to 4. **Figure 5-12(a)** shows the effect of block size on the performance of the 2-D sliding correlator for 2x2 sliding. As expected, the 16x16 macro-block gives better results.



(a)



(b)



(c)

Figure 5-11 Performance of the 2-D DCT watermarking scheme for several video sequences: **(a)** without sliding, **(b)** with 2x2 sliding and **(c)** comparison between these two cases.

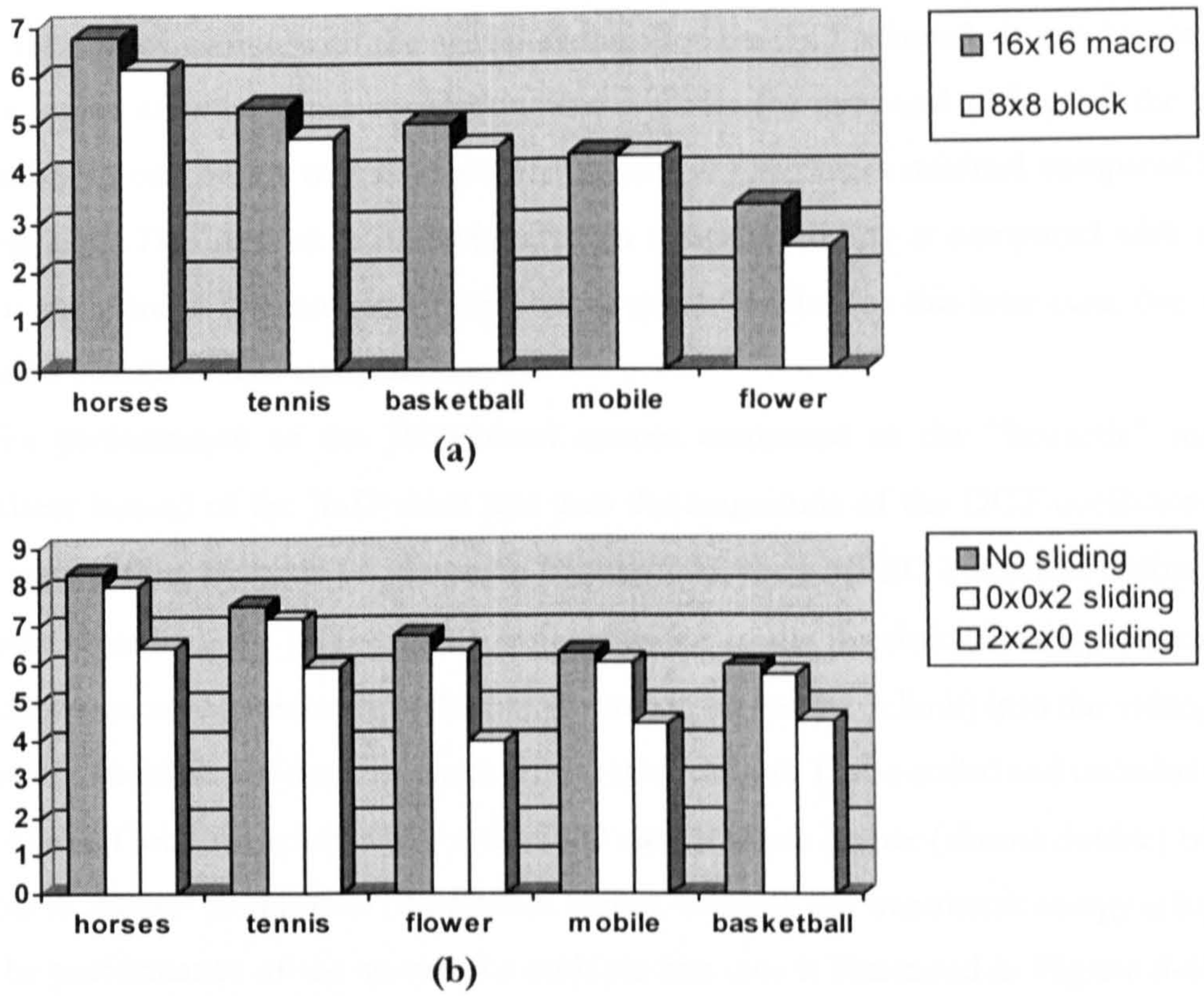


Figure 5-12 DCT domain watermarking: **(a)** the effect of block dimension on the 2-D sliding correlator, for 2x2 sliding and **(b)** the effect of spatial and temporal sliding on the performance of the 3-D system.

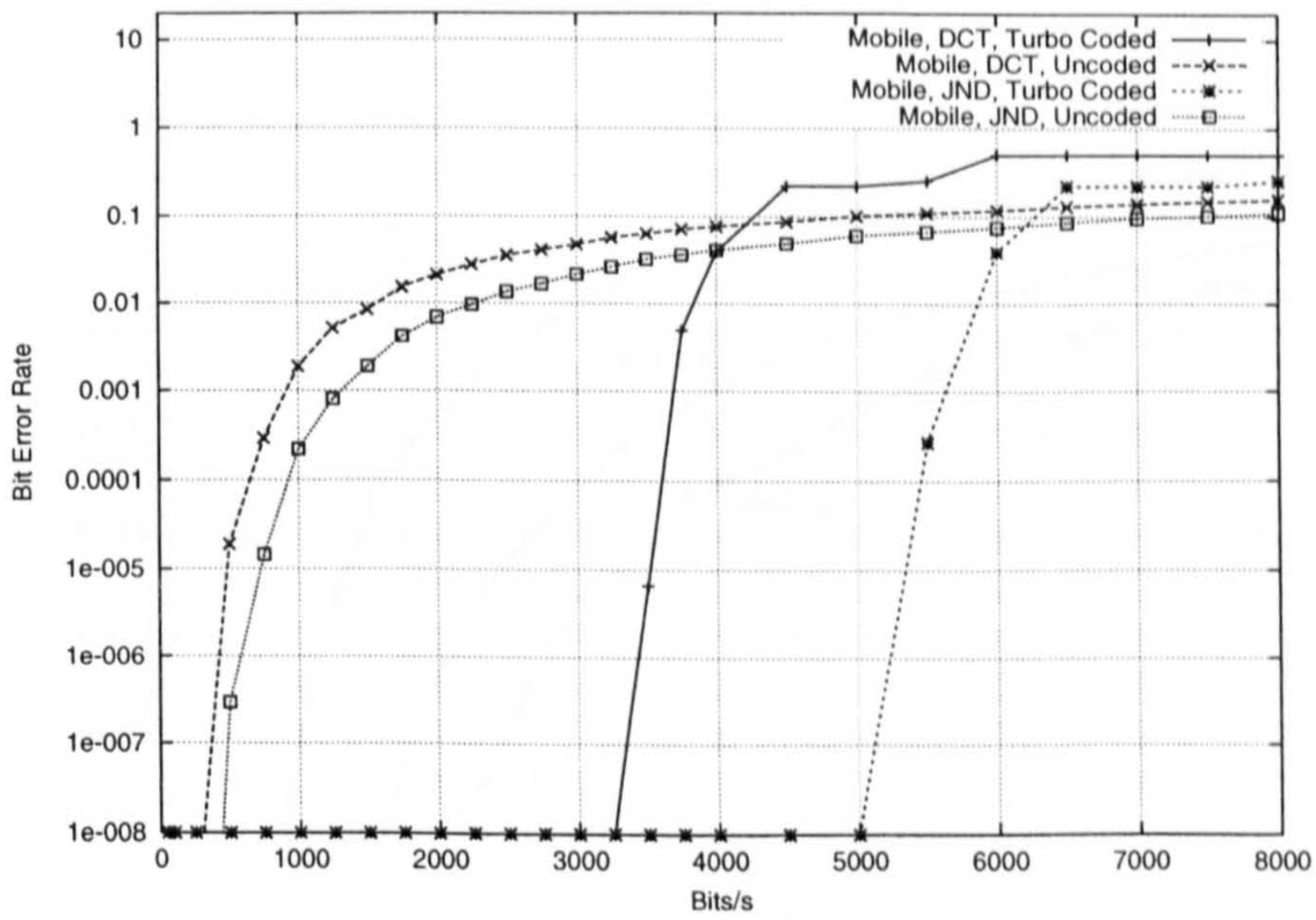


Figure 5-13 The capacity of the JND-based system compared with the "heuristic" marking, under 6Mbps MPEG2 attack.

For the same parameters of the scheme as in the 2-D case, **Figure 5-12(b)** compares the case of temporal sliding with the spatial sliding (for the 3-D scheme). It can be seen that due to the higher effective cross-correlation area available for temporal sliding (all the blocks corresponding to one bit are moved together) the loss in this case is minimal compared to the non sliding case. This difference is obvious when temporal sliding is compared with spatial sliding (for the same 3-D correlator). The SNR is much smaller for this later case, due to the much smaller effective cross-correlation area.

The performance of the JND-based system compared to the "heuristic" marking system (where instead of the JND value one uses the magnitude of the DCT coefficients and an appropriate scaling factor α) is shown in **Figure 5-13**, for a MPEG2 attack at 6Mbps. As it can be easily remarked, the JND scheme is net superior to the heuristic marking, since allows one to embed a much higher energy (close to the maximum possible limit) into the video, while maintaining the invisibility of the watermark. This is true for both Turbo coded and uncoded cases. The gain of the Turbo coded system for the JND case is much higher (almost double) because the channel is "better" in this case (the SNR is higher, because the watermark energy is higher).

The performance of the system for multiple line cuts is illustrated in **Figure 5-14**. The figure illustrates 3 distinct cases: first, only one line (line 288) is cut and the watermark is recovered without performing any sliding at al; in the second case, two lines are cut (line 192 and line 384) and the watermark is recovered still without any sliding at al; and finally, 2x2 spatial sliding is employed in order to recover the watermark from the attacked image.

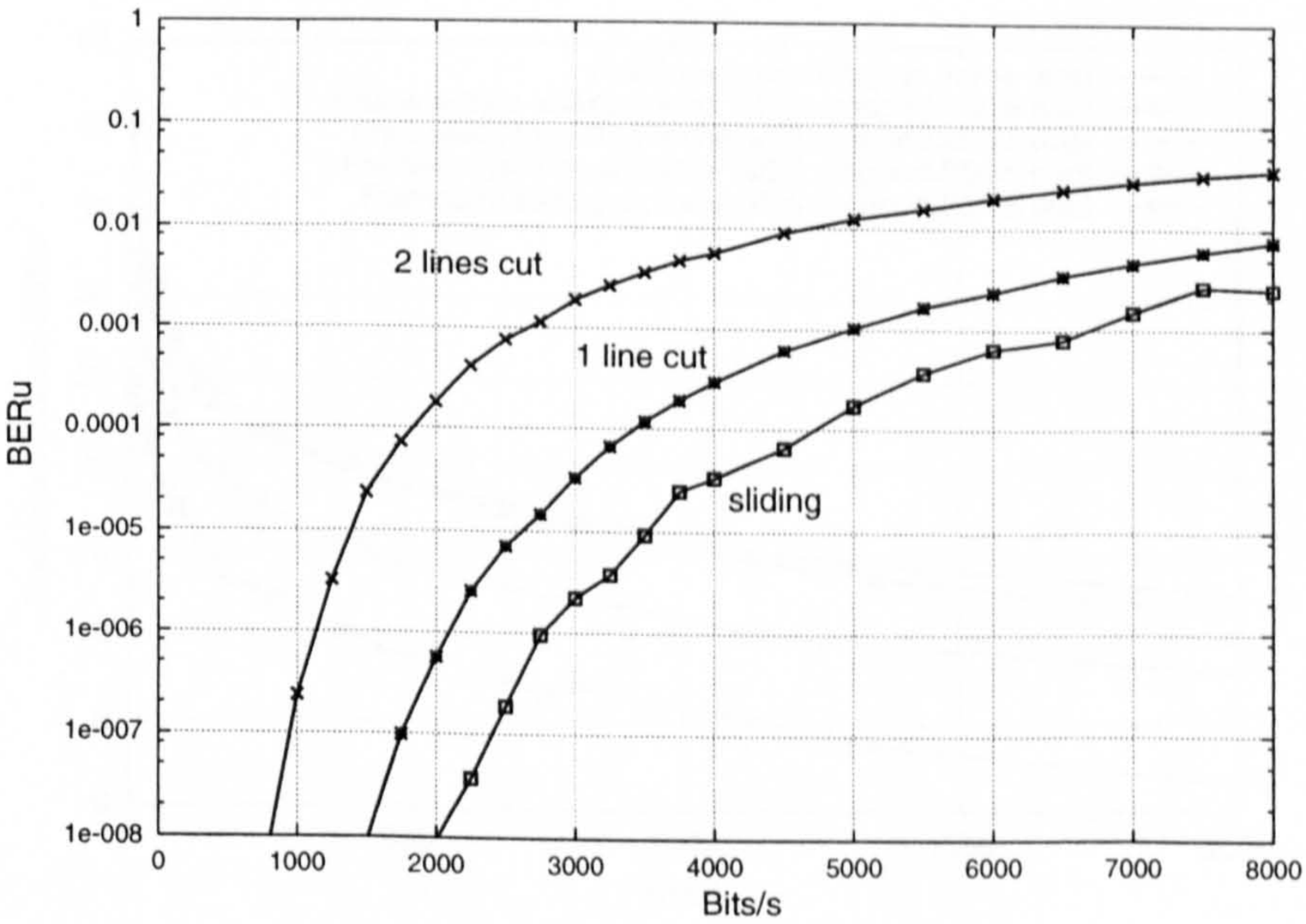


Figure 5-14 The performance of the JND-based system under multiple line cuts and the effect of sliding, for an uncoded system and typical video sequence "basketball".

Although the efficiency of the spatial sliding is not very high (when the video is not attacked the system's capacity is higher than 8000bps) even for the 3-D correlator, the capacity still improves significantly compared to the cases when no sliding is performed.

Using the same 3-D correlator, this time for frame cuts, leads to an entirely different situation. In **Figure 5-15** are illustrated again three distinct cases: first the frame number 5 (out of 25 frames) is cut and the watermark recovered without any temporal sliding; in the second case the frame number 10 is removed and again the watermark is recovered without any temporal sliding; finally temporal sliding is employed in order to recover both these attacks. One can easily see that cutting frame number 5 is a much worse attack than cutting frame number 10, because in this case the watermark is completely desynchronised starting with the frame number 6 rather than starting with the frame number 11. When temporal sliding is involved, the efficiency of the cross-correlator is very high, since moves together a much larger area compared to the case of temporal sliding. That's why the results for temporal sliding are very close to the un-attacked situation. In this case the marking depth was 2 frames.

The impact of the marking depth on the system's performance is illustrated in **Figure 5-16**, for an attack consisting in a line cut combined with 6Mbps MPEG2 compression. The diagram shows the results for a marking depth of 5 frames compared to the case of the "classical" 2-D correlator (which can be regarded as a 3-D correlator with a marking depth of one frame). The results are presented for both uncoded system and coded system, for a block length of the code N=128. As expected, the performance increases with the marking depth.

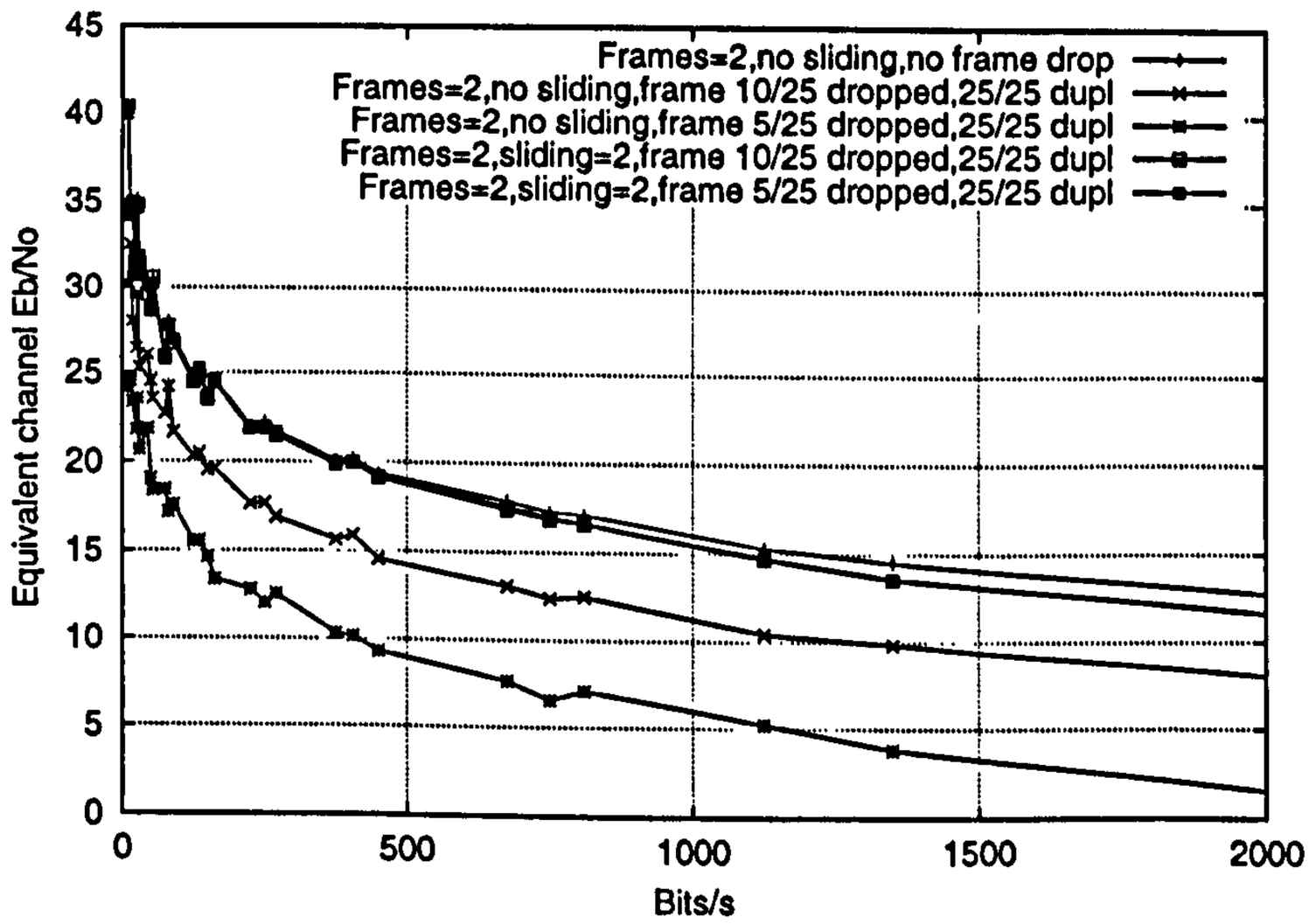


Figure 5-15 The performance of the JND-based system under frame cuts and the effect of sliding, for an uncoded system and typical video sequence "basketball".

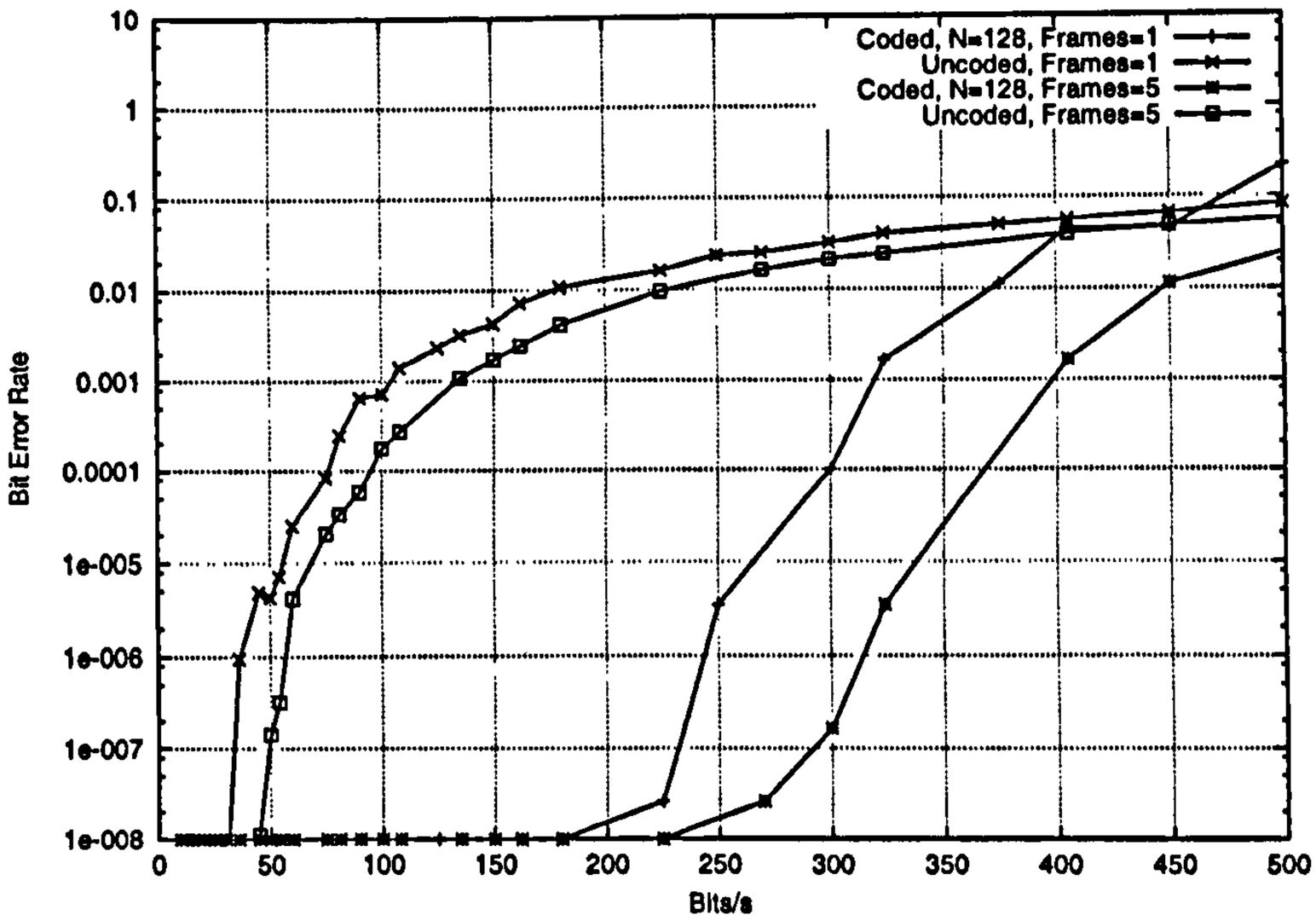


Figure 5-16 The influence of the marking depth on the system's performance under combined attack (line cut plus 6Mbps MPEG2 compression).

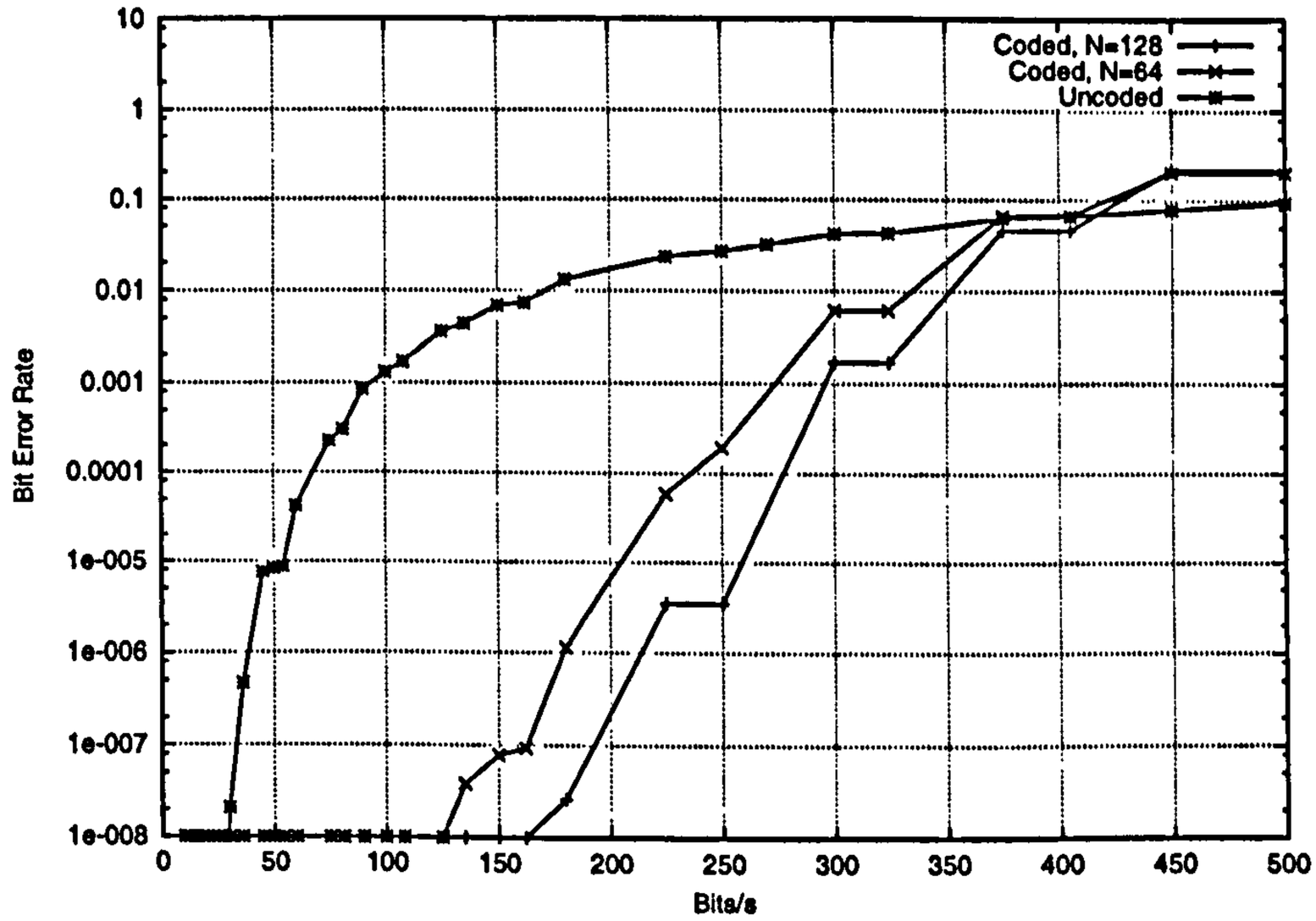


Figure 5-17 The influence of the Turbo code's block length on the system's performance under 3Mbps MPEG2 compression, for "basketball" video sequence.

The impact of the Turbo code and the importance of the block length of the code are presented in **Figure 5-17** for a 3Mbps MPEG2 compression attack. Using a Turbo code with a block length of 64 bits improves the capacity of the system about four times, while for a length of 128 bits the capacity is five times higher. This fact can be explained by the fact that the performance of the Turbo code increases with the length of the block (i.e. the interleaver length) as was shown in **Figure 4-7**.

One concern of the content providers is that several users could collude their watermarked material (each copy of the same video sequence contains a different watermark) in order to "remove" the watermark. This is done by adding a number of watermarked copies together and then taking their average as the attacked video; doing this with a sufficiently high number of copies, will "disable" the watermark. Of course the attacker doesn't have a large number of sequences, so the collusion should be ineffective for a reasonable number of copies. The results for this attack are presented in **Figure 5-18**, for the uncoded case and for a different number of copies colluded together. One can see that this attack is quite mild. Even without Turbo coding and when 5 sequences are colluded together, the capacity is still around 1000bps. Applying coding to the scheme will result in capacities larger than 8000bps.

Finally, another potentially damaging attack is the VCR attack. In this case the digital programme (video sequence) is recorded to an analogue tape using a standard VCR. The content provider wants to be able to recover the watermark even in this case.

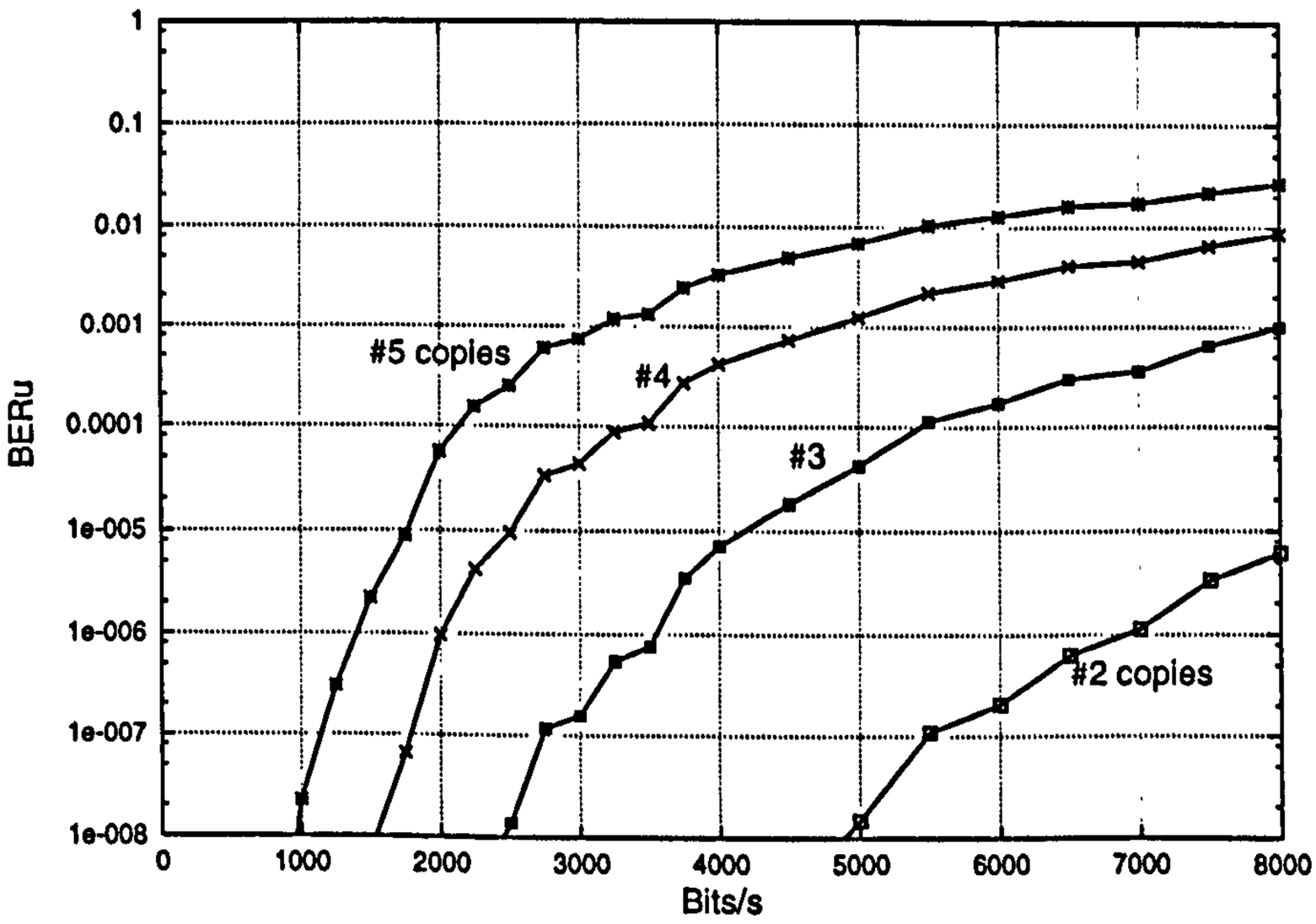


Figure 5-18 Collusion attack with a variable number of copies and its effect on system's performance.

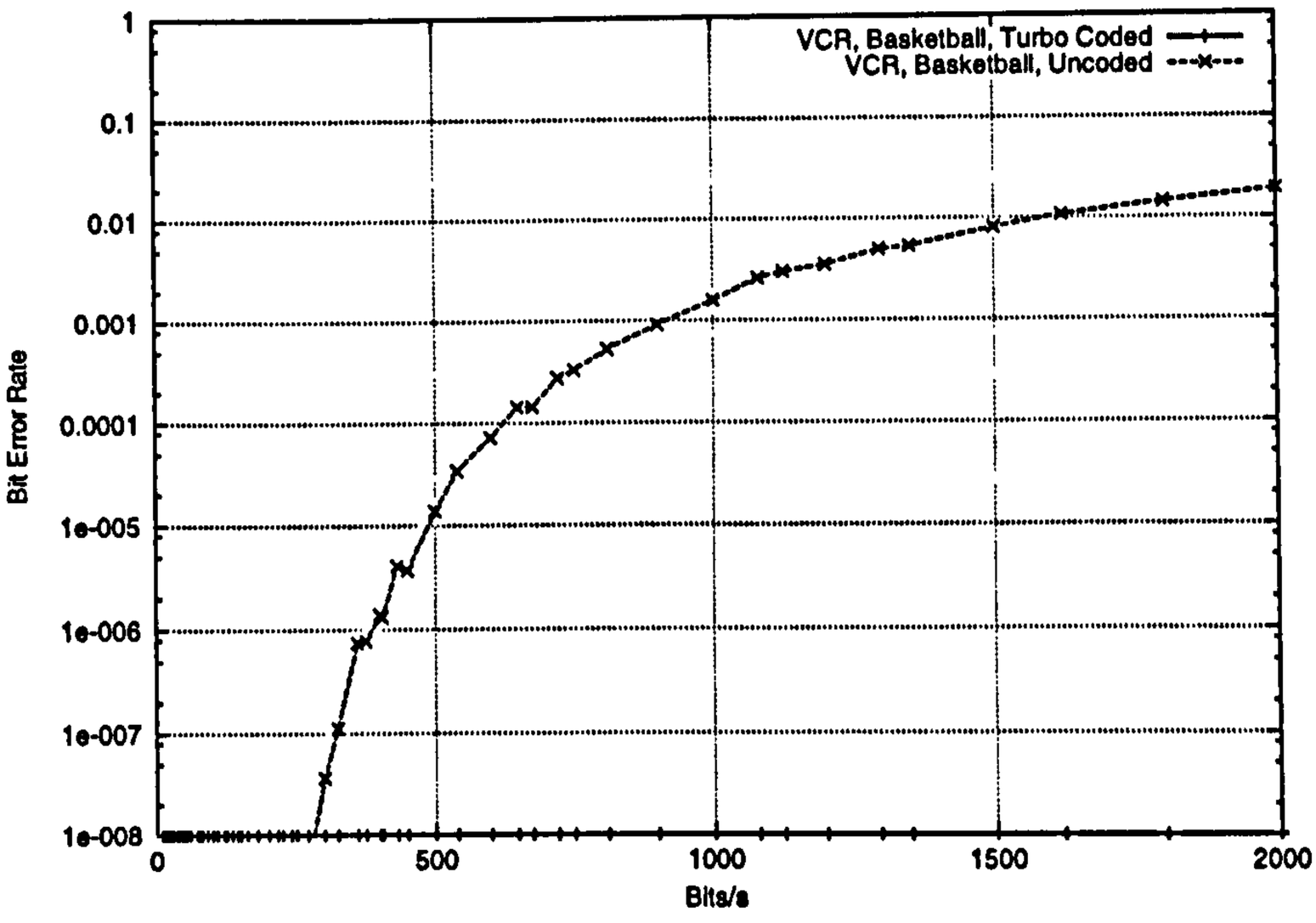


Figure 5-19 The VCR attack: the video is recorded on an analogue tape and then re-recorded in digital format using a specialised digital capture card.

Of course this can be done only after the analogue signal is transformed back to digital domain, for example by using a specialised video capture card. This attack affects the video in several ways: the signal is converted twice involving D/A ↔ A/D converters, the colour components could be slightly altered (not important in our case) and finally some jitter could be present due to the analogue recording process. In fact, this attack proves to be relatively mild as well, although is more damaging than the collusion attack. As **Figure 5-19** illustrates, even in the uncoded situation the capacity of the system is relatively large, around 350 bps. By using Turbo coding the capacity of the system exceeds 2000bps.

5.6 Conclusions

The frequency marking domain is known to give better results compared to the spatial domain watermarking techniques. The DCT in particular has also the advantage of being widely used in image processing, especially for compression. Many HVS models were developed in this context and there is relatively easy to adapt such a visual model to the requirements of watermarking.

Following this idea, this chapter presented one of the best HVS models available: the JND model and its application to the watermarking framework. This proves to be a very successful step in improving the watermarking system. With its highly adaptive nature, the JND model leads to a maximal robustness and maximal capacity watermarking system, while still preserving the invisibility of the watermark. As section 5.5 proved, the JND model almost doubles the capacity of the system.

Applying communication theory to watermarking becomes a more and more popular choice in watermarking community. By seeing the watermark channel as a communication channel, one could employ error correction codes to protect the watermark. Turbo codes are one of the best candidates for such a system, as already discussed in Chapter 4. The capacity of the watermarking system increases up to 4-5 times under 3Mbps MPEG2 compression attack, when Turbo codes are employed.

Although not as successful as the previously described methods, the system can be improved further by increasing the effective (local) cross-correlation area, using the macro-block concept and extending the system to account for the temporal dimension. Furthermore this extends the capability of the system to counteract time sync errors like frame dropping. The temporal sliding correlator proves itself to be highly efficient in combating frame dropping while achieving minimal loss compared to the non attacked case. Spatial shifts and line/column cuts can be handled without major problems. Alternative modulation techniques were also investigated leading to almost similar results (just slightly worse).

Cropping can be handled as well, giving quite good results. For example, even with an extreme attack like cropping a small 200x200 region from the video sequence, the system gives a capacity of some 1250bps (**Figure 6-9(a)**).

Attacks like collusion and VCR are relatively mild attacks and they are not posing a real threat to the system. Even without Turbo coding, the system can still achieve reasonable capacities, in fact much higher than the one required for broadcast monitoring [Cheveau et al, 2000].

Probably the most fearsome attack remains MPEG2 compression. While for a low compression at 6Mbps the system performs very well, giving a capacity of some 5000bps, this quickly drops to some 150bps for MPEG2 compression at 3Mbps while at 2Mbps the watermark cannot be retrieved at all, being completely lost.

Chapter 6

"All animals are equal but some animals are more equal than others."

George Orwell, 1903-1950, "Animal Farm"

Wavelet Domain Watermarking

This chapter begins with an introduction to the wavelet transform, with the accent on the 2-D DWT (Discrete Wavelet Transform) case. The multiple advantages of the DWT transform are discussed and compared with the traditional FFT/DCT transforms [Jain, 1989], [Misiti et al, 2001], taking into account the specific framework of digital watermarking. Choosing a proper basis constitutes an important step which will be also discussed.

Due to major advantages of the DWT, the wavelet coefficients are one of the most suitable places to insert a watermark. The proposed watermarking system is described in detail during this chapter, including the HVS aspects of the scheme and error correction. The performance of the system will be then analysed for both image watermarking (in order to compare the results with the existing image watermarking schemes described in the literature) and video watermarking.

6.1 Short introduction to the Wavelet transform

Historically speaking the wavelet analysis is a relatively new method, although some of the mathematical background dates back to the theory of Fourier in the nineteenth century. Fourier set the basis of the frequency analysis which for a long time was the best and the only approach existent in signal analysis.

The research gradually moved from frequency-based analysis to scale-based analysis when the researchers realised that an approach measuring average fluctuations at different scales might prove less sensitive to noise. And so, the wavelet transform was born. The first recorded mention of what we call now a "wavelet" dates back to 1909 in Alfred Haar's thesis. The concept of wavelets in its present theoretical form was proposed later by Jean Morlet. The methods of wavelet analysis have been developed mainly by Yves Meyer and his colleagues, and the main algorithm was provided by Stephane Mallat in 1988. Since then, the research has become international [Burke-Hubbard, 1998], [Misiti et al, 2001].

6.1.1 Wavelet versus Fourier

Fourier analysis

The Fourier transform [Jain, 1989], [Misiti et al, 2001] is perhaps the most well-known way of analysing a signal. The signal is break down into its constituent sinusoids of different frequency e.g. transforming the signal from the time domain to frequency domain.

This analysis is very useful since in most of the cases the frequency content of a signal is very important. One important drawback of the Fourier analysis is that during this transformation time information is lost. In other words looking at the Fourier transform of a signal is impossible to say when a particular event took place. For stationary signals this is not a problem, but for real life signals which usually are non-stationary the transitory events are of capital importance.

Short-time Fourier analysis

In a bid to overcome this deficiency, the signal can be windowed e.g. the Fourier transform analyses only a small section of the signal at a time, given by the size of the window. Introduced by Gabor in 1946, this technique is called Short-Time Fourier Transform (STFT) [Misiti et al, 2001].

In this case the signal is mapped into a two-dimensional function of time and frequency. The STFT offers some information about when and at what frequencies a signal event occurs, but only with limited precision, given by the size of the window.

This compromise between time and frequency information is very useful but has a drawback too: once the size of the window is chosen, that window size is used for all frequencies. The STFT is a good start, but many signals require an even more flexible

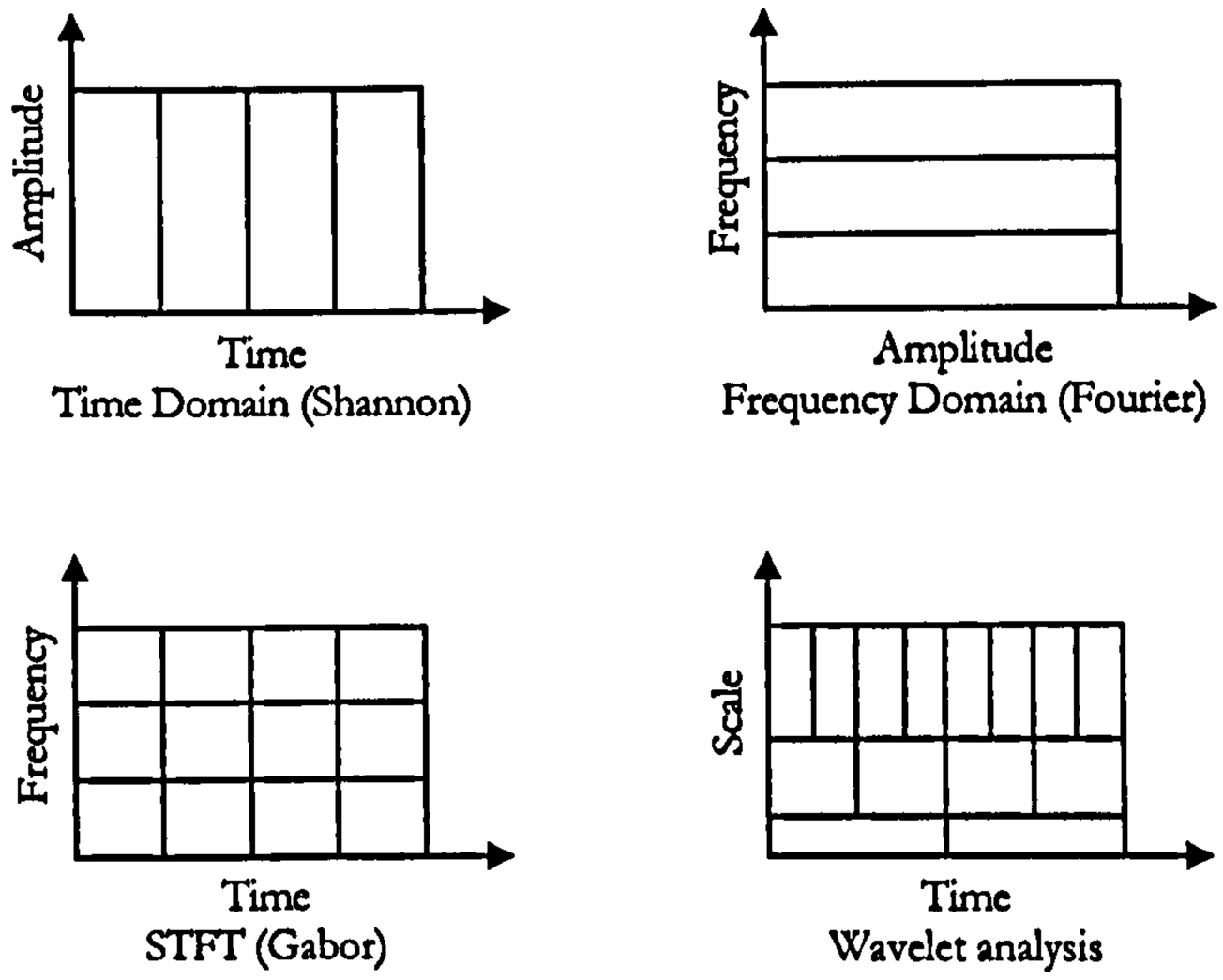


Figure 6-1 Methods of signal analysis: a comparison between time domain, Fourier, STFT and wavelet analysis

approach, where the window size can be varied in order to determine more accurately either time or frequency.

Wavelet analysis

The later requirement can be achieved by the wavelet analysis, which offers a windowing technique with variable sized regions. The wavelet transform allows the use of long time intervals where we want more precise low-frequency information and shorter intervals where we want high frequency information.

The concept is illustrated in **Figure 6-1** compared with the other 3 traditional approaches: time domain analysis, Fourier analysis and STFT analysis.

One major advantage of the wavelet transform is the ability to analyse a localised area of a larger signal. In this way, wavelet analysis is capable to reveal aspects of the data that other signal analysis techniques miss, for example breakdown points, trends, discontinuities in higher derivatives and self-similarity.

6.1.2 Wavelet transform

Unlike the FFT and the DCT, the DWT is a hierarchical transform which offers the possibility of analysing a signal at λ different resolutions or levels (with λ integer). Such multiresolution analysis gives a frequency domain representation as a function of time (or space in the 2-D case) i.e. both time/space and frequency localisation as shown in **Figure 6-1**.

One cannot achieve infinite resolution simultaneously in both time and frequency (Heisenberg uncertainty principle); high time resolution forces poor frequency resolution and vice-versa. This trade-off is used by the wavelet transform to provide multiresolution analysis.

In order to analyse the signal in terms of both frequency and time, the analysing functions must have different frequencies and they also have to be localised in time. Formally we refer to scale and resolution, where, for the dyadic case, scale is defined as $a = 2^{\lambda}$ and resolution as $r = \frac{1}{a} = 2^{-\lambda}$. Scale in this case means simply stretching (or compressing) the wavelet. The smaller the scale factor, the more compressed is the wavelet, or in other words, the greater the resolution, the smaller and finer are the details that can be analysed. There is a correspondence between wavelet scales and frequency as revealed by wavelet analysis:

- Low scale → Compressed wavelet → Rapidly changing details → High frequency
- High scale → Stretched wavelet → Slowly changing, coarse features → Low frequency

A wavelet can be defined as a waveform of effectively limited duration that has an average value of zero as opposed of the infinite duration sine waves used in the Fourier analysis. Also the wavelets tend to be irregular and asymmetric rather than smooth and predictable as the sine waves. A representation of the Antonini 7.9 wavelet for different scales is given in **Figure 6-2.**

For the 1-D case, a certain wavelet is defined by the mother wavelet function $\Psi(x)$ and a scaling function (or father wavelet) $\Phi(x)$, and the analysing wavelets are scaled

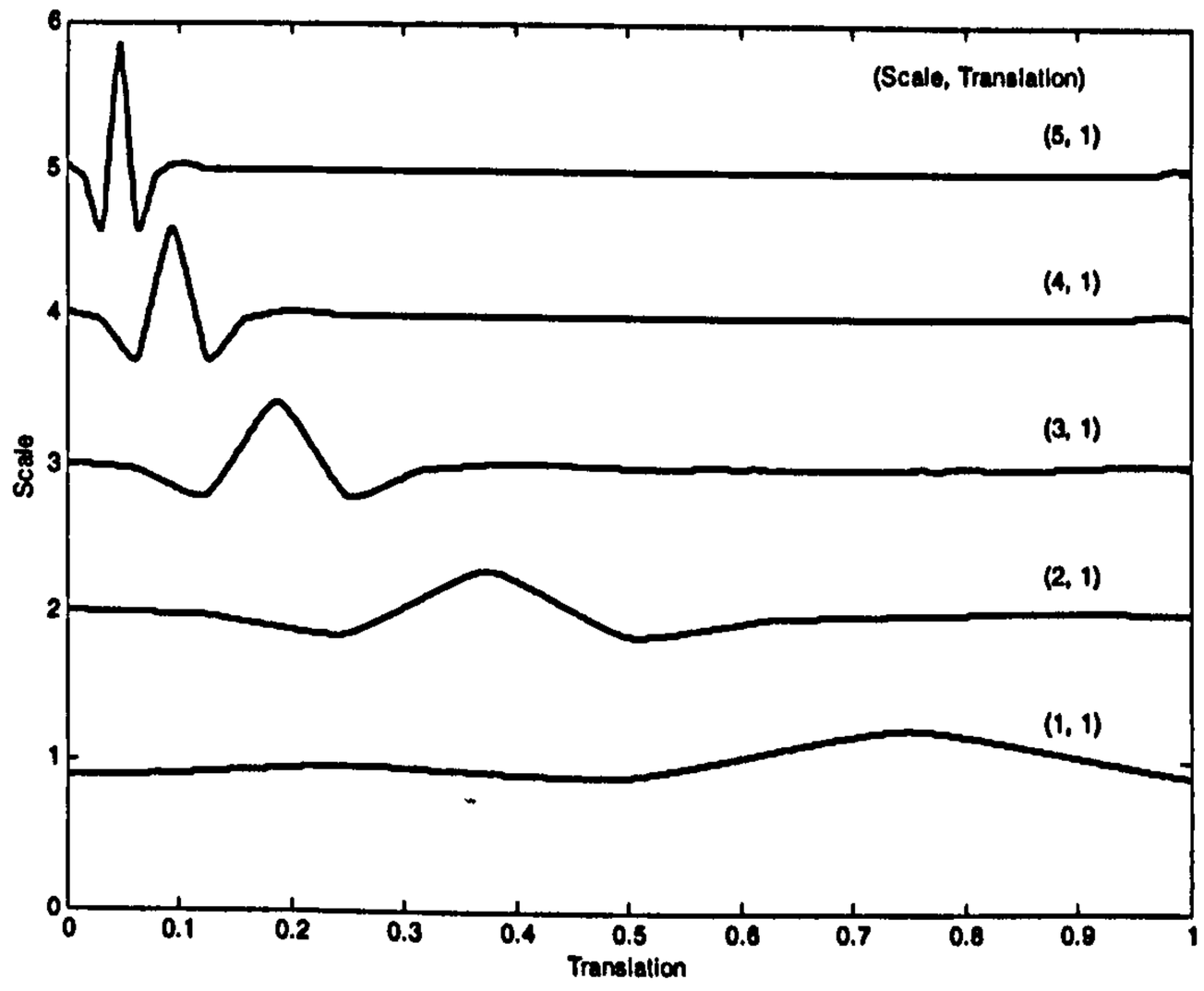


Figure 6-2 The Antonini 7.9 wavelets at various scales (same translation).

and translated versions of the mother wavelet:

$$\frac{1}{\sqrt{a}}\Psi\left(\frac{x-b}{a}\right) \tag{6.1}$$

The wavelet analysis breaks up the signal into shifted (translated) and scaled versions of the original (mother) wavelet. Defining translation $b = ka$, (where k, λ are integers) the dyadic case becomes:

$$\begin{aligned}\Psi_{\lambda,k}(x) &= 2^{-\frac{\lambda}{2}}\Psi(2^{-\lambda}x-k)\\ \Phi_{\lambda,k}(x) &= 2^{-\frac{\lambda}{2}}\Phi(2^{-\lambda}x-k)\end{aligned} \tag{6.2}$$

Given the input signal $f(x)$, a wavelet coefficient can be defined as:

$$C(\lambda,k) = \int_{-\infty}^{\infty} f(x)\Psi_{\lambda,k}(x)\,dx \tag{6.3}$$

For the 2-D case, the mother wavelet can be described as:

$$\frac{1}{\sqrt{a_1 a_2}}\Psi\left(\frac{x-b_1}{a_1},\frac{y-b_2}{a_2}\right) \tag{6.4}$$

In this case (b_1, b_2) represents the translation vector and (a_1, a_2) is the scaling parameter. Furthermore there are one scaling function $\Phi(x,y)$ and three wavelet functions $\Psi_\theta(x,y)$, where θ denotes orientation:

$$\Phi(x,y) = \Phi(x)\Phi(y) \tag{6.5}$$

and

$$\begin{aligned}\Psi_H(x,y) &= \Phi(x)\Psi(y)\\ \Psi_V(x,y) &= \Psi(x)\Phi(y)\\ \Psi_D(x,y) &= \Psi(x)\Psi(y)\end{aligned} \tag{6.6}$$

Different orientations extract different features of the frame, such as vertical, horizontal, and diagonal information. This fact is very well illustrated in **Figure 6-3**. Generally speaking, edges and textures will be represented by large coefficients in the high frequency sub-bands, and are well localised within the sub-band.

In practice wavelet analysis is performed using multilevel filter banks. Essentially this comprises a succession of filtering (lpf and hpf) and sub-sampling operations and has been widely described in the literature [Kingsbury, 1997], [Antonini et al, 1992], [Villasenor et al, 1995], [Watson et al, 1996 and 1997] and [Xia et al, 1998].

6.1.3 Main applications of the Wavelet transform

Several thousand papers have been written within the last 15 years about the wavelets and their applications. This proves once more the success of wavelet analysis.

Some of the most popular applications are briefly presented below. Many applications were developed for signal/image processing, including de-noising and compression. Watermarking tends to become another successful application area. Probably one of the most popular applications of the wavelets is the compression of the FBI fingerprints and the recent JPEG 2000 standard.

Medicine is a very prolific application field for wavelets, especially in heart diagnosis (EKG/ECG – Electrocardiography), EEG (Electroencephalography), mammography and MRS (Magnetic Resonance Spectra).

Many papers were published in oceanography and earth studies.

6.2 Choosing the right basis

For watermarking one needs to select an appropriate wavelet or basis. Most of the basis development has taken place in the context of image compression [Villasenor et al, 1995], and fortunately watermarking and compression have many things in common. On the other hand, it is very important to choose a basis that offers compact support. The smaller the support of

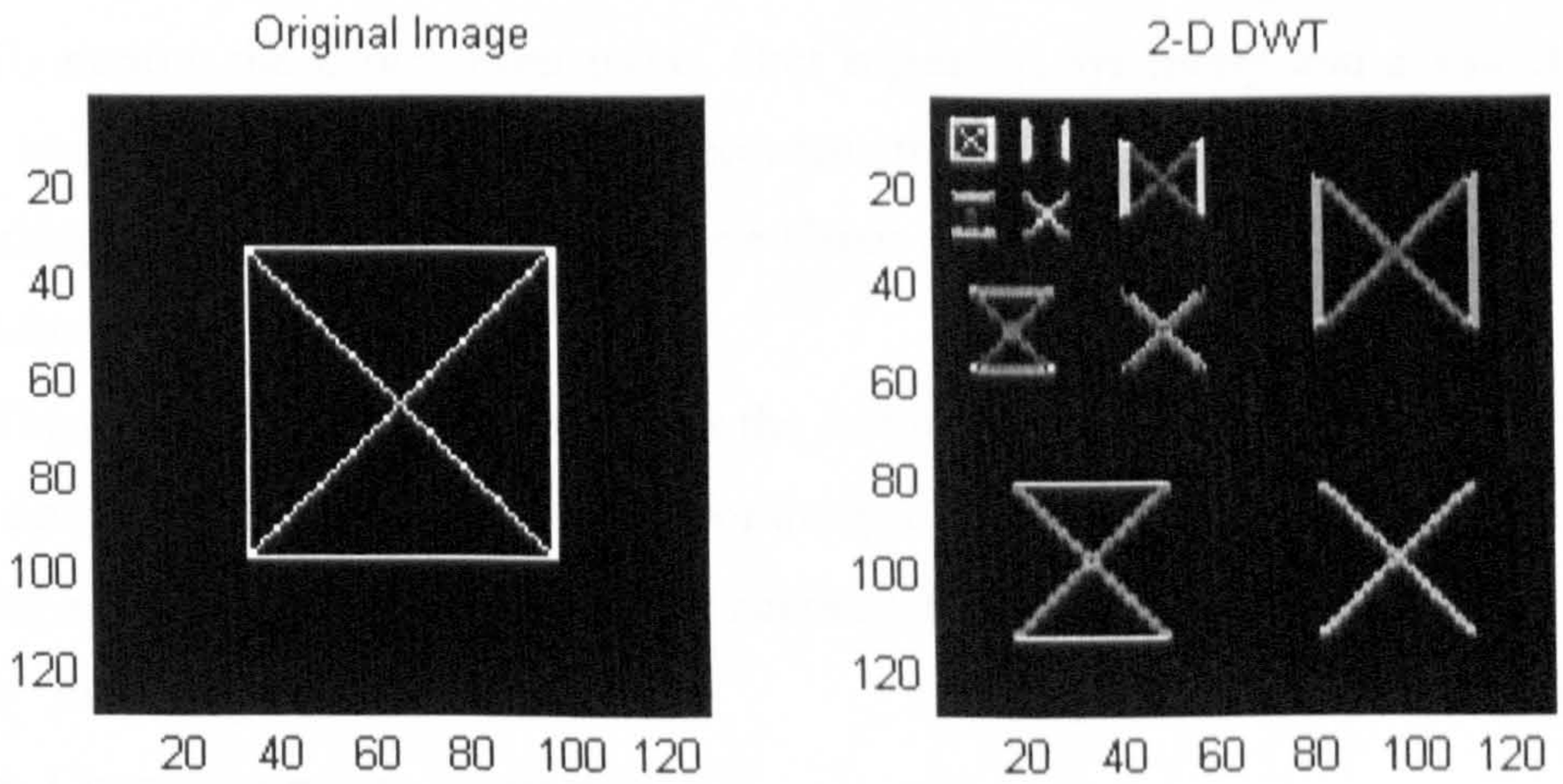


Figure 6-3 The 2-dimensional DWT: the original image and the wavelet decomposition for $\lambda = 3$.

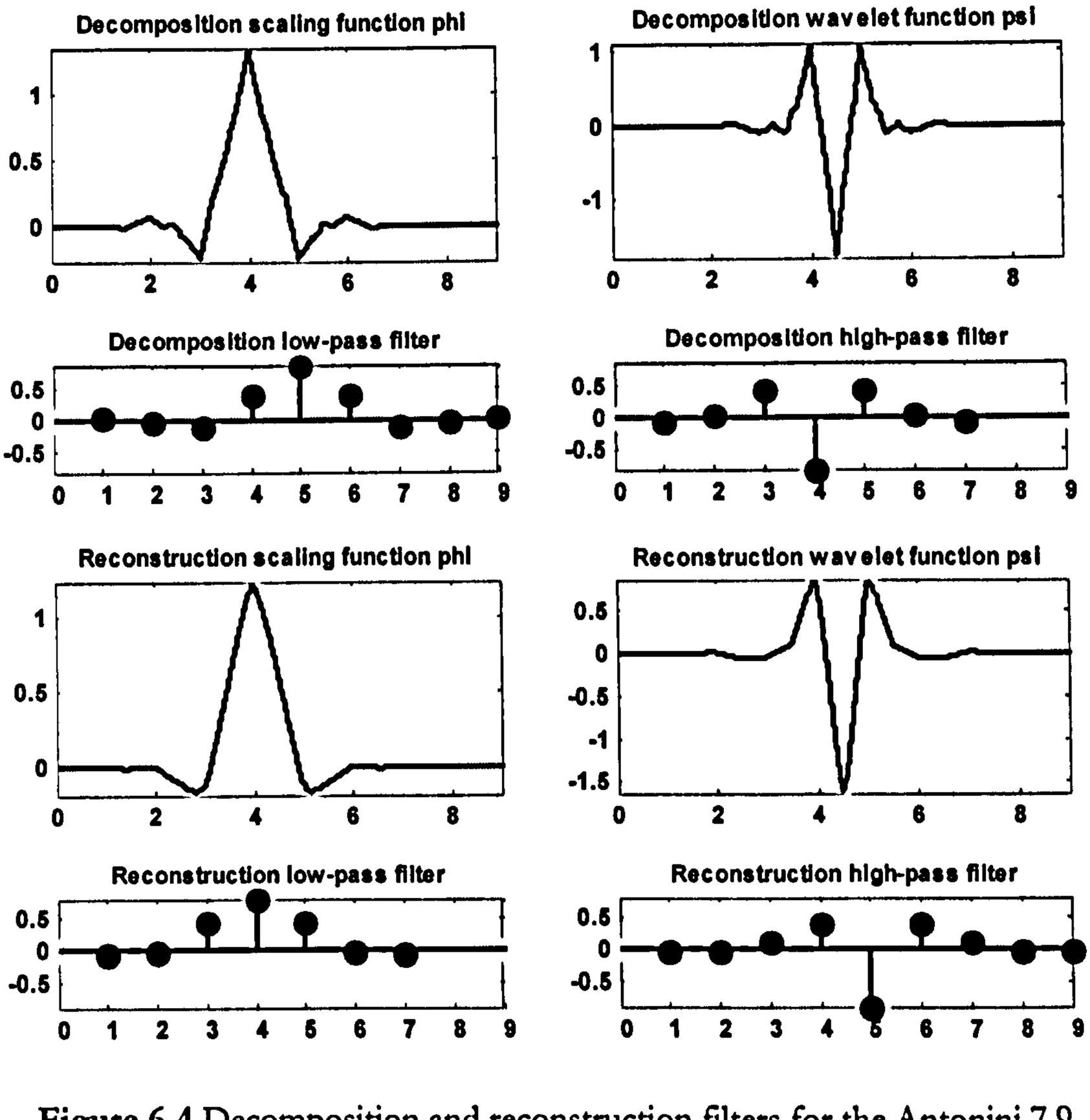


Figure 6-4 Decomposition and reconstruction filters for the Antonini 7.9 wavelet and the corresponding wave shapes.

the wavelet, the less nonzero wavelet coefficients will correspond to an edge for example, so basically the transform compacts more energy in the high frequency sub-bands [Lewis et al, 1992]. Also we are restricted to a class of either orthogonal or bi-orthogonal wavelets.

To narrow the choice even more, filter regularity, symmetry and a smooth wavelet function are important aspects for the reconstructed image quality. In addition, from the watermarking perspective this time, the selected basis must have a reasonably good HVS model designed for it.

The wavelet selected for this work is the Antonini 7.9 wavelet, presented in **Figure 6-4**, which is one of the best wavelets available for image compression [Kingsbury, 1997], [Antonini et al, 1992] and [Villasenor et al, 1995]. Its main properties are highlighted below:

- Bi-orthogonal wavelet
- Compact support, symmetric
- Good regularity (each filter has 2 factors)
- The lpf and hpf are quite similar

- Simple filters (only 7 and respectively 9 taps)
- Linear (zero) phase
- HVS model available [Watson et al, 1996 and 1997]
- Smooth wavelet function

This wavelet is widely used in image compression algorithms like EZW and SPIHT, but perhaps its most important application to date is the FBI fingerprint compression standard which uses this particular basis.

6.3 Advantages of the wavelet transform

The basis function for the DFT ($f(x)=\exp(i\omega x)$) or DCT (infinite cosine) has perfect localisation in frequency but is not time/space localised. In contrast, as already mentioned, wavelets offer a trade-off between time/space and frequency/scale, and so a watermarking scheme based on the DWT will produce a watermark with both spatially local and spatially global support (see **Figure 6-3**). This localisation makes a wavelet based scheme more robust than the DCT scheme, given geometric attacks such as cropping and scaling.

For instance, in the case of cropping, the lower frequency levels will be affected more than the high frequency ones, because of the fact that the watermark from the higher levels corresponds to a smaller spatial support. Looked at in the frequency domain, cropping corresponds to convolving the frequency components with a *sinc* function, where the width of the main lobe is inversely proportional to the width of the cropped window size [Podilchuck et al, 1998]. This will affect all the frequency components of any scheme based on a global transform, but since the wavelet scheme has a watermark with local spatial support, the watermark will be unaffected by the cropping.

For scaling, because the DWT coefficients are localised both in space and frequency,

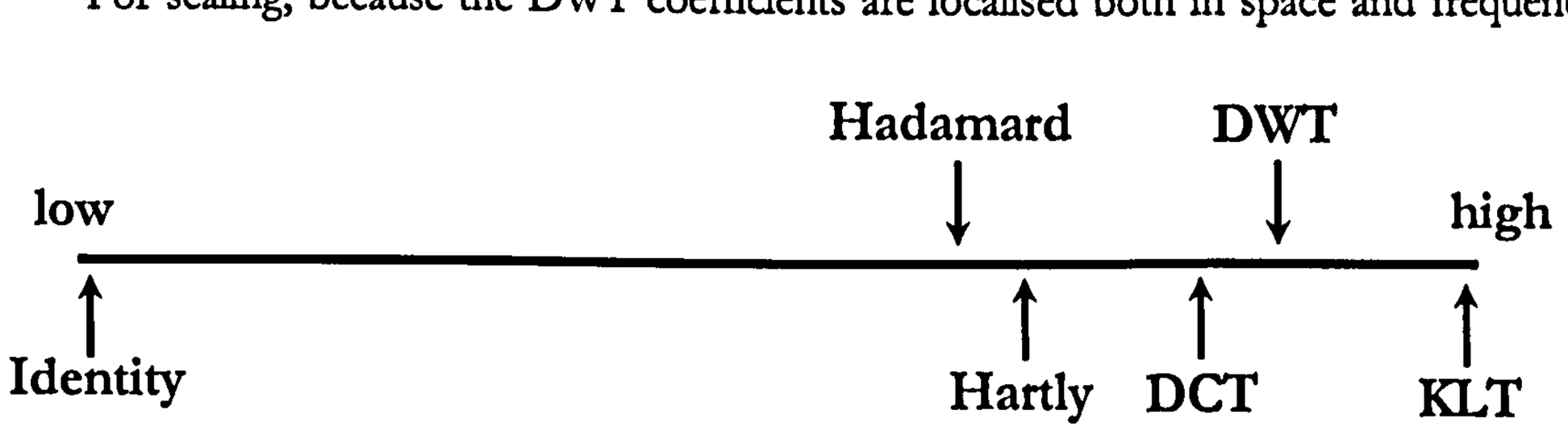


Figure 6-5 The energy compaction scale for several transforms.

whilst the DCT coefficients are only localised in frequency, it is likely that this kind of attack will be less serious for a DWT scheme. Simulation confirms this to be the case. Finally, the global spatial support of a DWT scheme will tend to be robust to operations such as low pass filtering/compression (which attenuate high frequency levels).

Another fundamental advantage of the DWT lies in the fact that it performs an analysis similar to that of the HVS. In fact the wavelet transform can be regarded as a rough HVS model by itself. The HVS splits an image into several frequency bands and processes them independently. In a similar way, the DWT permits the independent processing of different sub-bands without significant perceptible interaction between them. Again, this is because the analysing functions Ψ are localised in space, being zero outside a space domain U i.e. the signal values located outside of domain U are not influencing the values of the coefficients within U. Similarly, if Ψ is translated to position b, the wavelet coefficient will analyse the signal around b. This local analysis is specific to the compact support wavelets. Basically, for a small scale a local analysis is performed, whilst a large scale corresponds to a global analysis. **Figure 6-2** illustrates how the wavelet functions change for different scales.

Finally, more general advantages of the DWT are:

- It is not a block based transform, and so, the annoying blocking artefacts associated with the DCT are absent.
- Its hierarchical, multiresolution property offers more degrees of freedom compared with the DCT (for example separate or hierarchical cross-correlation).
- Higher watermark capacity and better robustness to attacks compared with the more traditional transforms (FFT, DCT).
- Lower computational cost than the FFT or DCT: $O(n)$ instead of $O(n\log(n))$, where n is the order of the transform input vector (lower hardware requirements) [Lumini et al, 2000], [Jain, 1989].
- Better energy compaction than both the FFT and DCT, in the sense that it is closer to the optimal Karhunen-Love transform as **Figure 6-5** depicts [Ramkumar et al, 1998-2]. More details about these transforms can be found in [Jain, 1989] and [Ramkumar et al, 1998-2].
- Some wavelets have useful invariance properties (for example complex wavelets are shift invariant [Kingsbury, 1997, 1998 and 1999]).

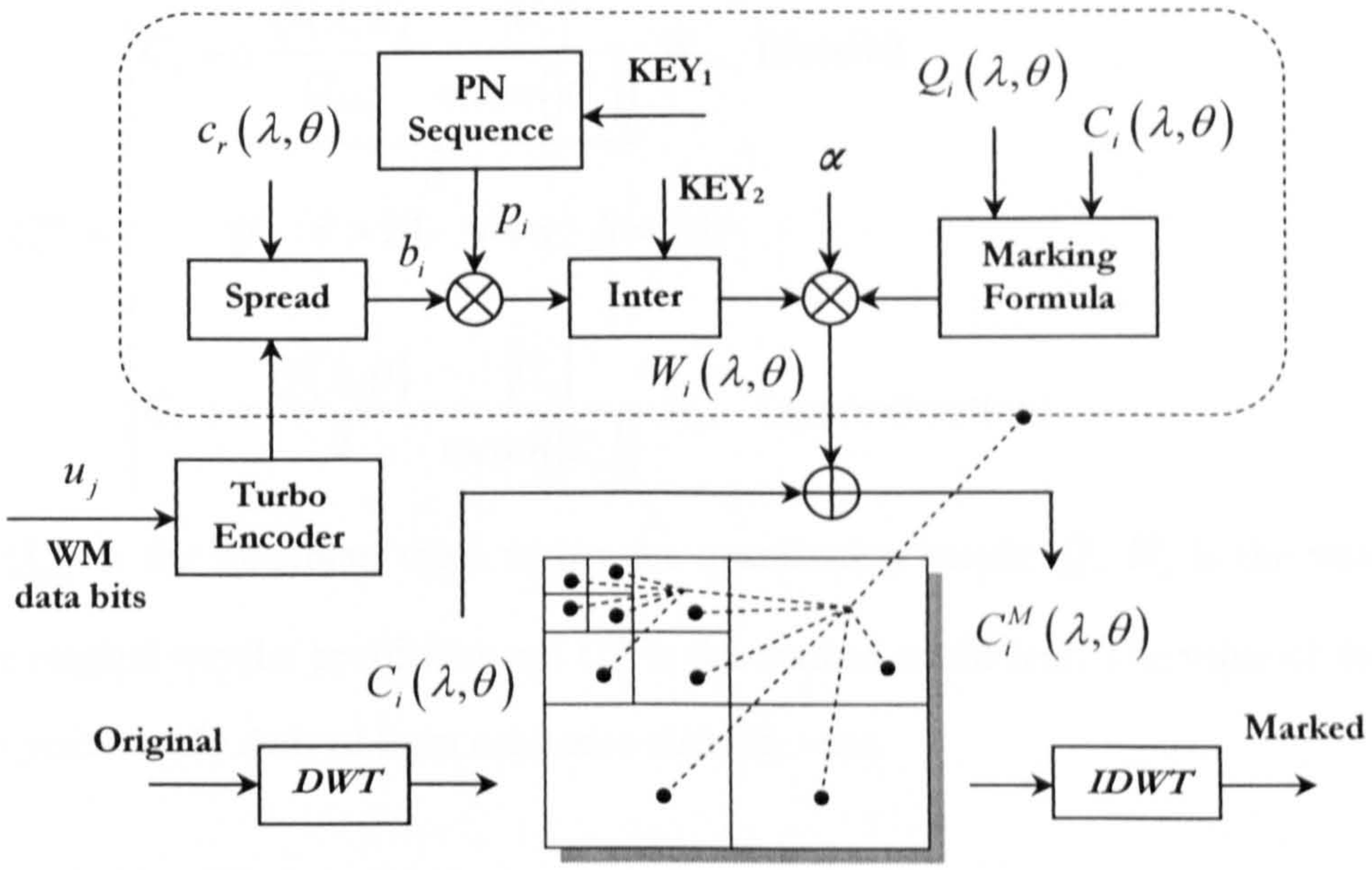


Figure 6-6 Wavelet-based watermark embedding.

6.4 The DWT-based watermarking scheme

6.4.1 Wavelet-based watermark embedding

Watermark embedding is illustrated in **Figure 6-6**. The preferred choice of the parameter λ is 3 (three levels of decomposition) [Misiti et al, 2001].

As for DCT systems, embedding uses the spread-spectrum approach and retrieval is via cross-correlation (matched filtering). The interleaver uses a separate key to that of the PN sequence in order to enhance system security and provide a random distribution of the data bits within each sub-band.

The hierarchical nature of the DWT is exploited here, by choosing to insert a self-contained watermark in each sub-band. This means that all of the data bits are inserted in each sub-band, the chip rate reducing as λ increases. Although reducing chip rate may appear to be a disadvantage, the advantage of this type of marking comes at the retrieval.

The watermark is embedded using amplitude modulation as follows:

$$C_i^M = \begin{cases} C_i + \underbrace{\alpha \frac{Q(\lambda,\theta)}{Q_{\min}} \cdot \frac{|C_i|}{mean(|C_i|)}}_{S} \cdot W_i, & \text{(details)} \\ \quad if \quad S > 24, \quad then \quad S = 24. \\ C_i + \alpha \frac{Q(\lambda,\theta)}{2} \cdot \frac{|C_i|}{mean(|C_i|)} \cdot W_i, & \text{(approximation)} \end{cases} \tag{6.7}$$

where $Q_{\min}$ is the minimum value within the quantisation matrix Q, W_i is the watermark, C_i is the original wavelet coefficient and C_i^M is the marked coefficient. The value of the factor S was experimentally derived from subjective visibility tests.

6.4.2 The HVS model

The HVS is incorporated in the quantisation matrix $Q(\lambda,\theta)$, where θ represents the orientation. Although this is a much simpler HVS model compared with the one used in the DCT scheme, the overall performance of the scheme is better. This illustrates once more the superiority of the wavelet transform over conventional transforms like FFT and DCT.

The matrix $Q(\lambda,\theta)$ offers only one quantisation factor for an entire sub-band, and incorporates only limited information about the HVS (essentially only the frequency sensitivity of the eye). In other words, the model is HVS dependent since it incorporates some aspects of the human vision (MTF of the eye), but unfortunately it is not media dependent, which constitutes an important drawback.

The quantisation matrix is computed according to the visual model developed by Watson for the Antonini 7.9 DWT [Watson et al, 1996 and 1997]:

$$\log Y = \log a + k\left(\log f - \log g_\theta f_0\right)^2 \tag{6.8}$$

where:

$$a = 0.495$$
$$k = 0.466$$
$$f_0 = 0.401$$
$$g_\theta = \{1.501,\ 1,\ 0.534,\ 1\},\ for\ \ \theta \in \overline{1\ldots4}$$

The rest of the parameters can be defined as:

$$f = r \cdot 2^{-\lambda} \quad \text{(cycles/degree)} \tag{6.9}$$

and:

$$r = dv \tan\left(\frac{\pi}{180}\right) \simeq dv \frac{\pi}{180} \tag{6.10}$$

where d is the display resolution in pixels/cm and v is the viewing distance in cm.

Finally, the quantisation factor for each sub-band is derived as:

$$Q(\lambda,\theta) = \frac{2Y_{\lambda,\theta}}{A_{\lambda,\theta}} = \frac{2}{A_{\lambda,\theta}} \cdot a \cdot 10^{k\left(\log \frac{2^{\lambda} f_o \, g_\theta}{r}\right)^2} \tag{6.11}$$

where $A_{\lambda,\theta}$ represents the basis function amplitude for the Antonini 7.9 wavelet [Watson et al, 1996 and 1997].

The quantisation matrix $Q(\lambda,\theta)$ is a rough measure of the visibility for each sub-band, and, as stated, it is not media dependent. This dependence is required for a robust watermark and is provided by the embedding algorithm in equation (6.7). This marks more heavily the high frequency sub-bands and the largest coefficients, since modification of these coefficients is less likely to incur visible artefacts.

6.4.3 Wavelet-based watermark recovery

Watermark retrieval is shown in **Figure 6-7**. As mentioned, it is advantageous to have a self-contained watermark (all data bits) in each sub-band, since a SNR can be determined for

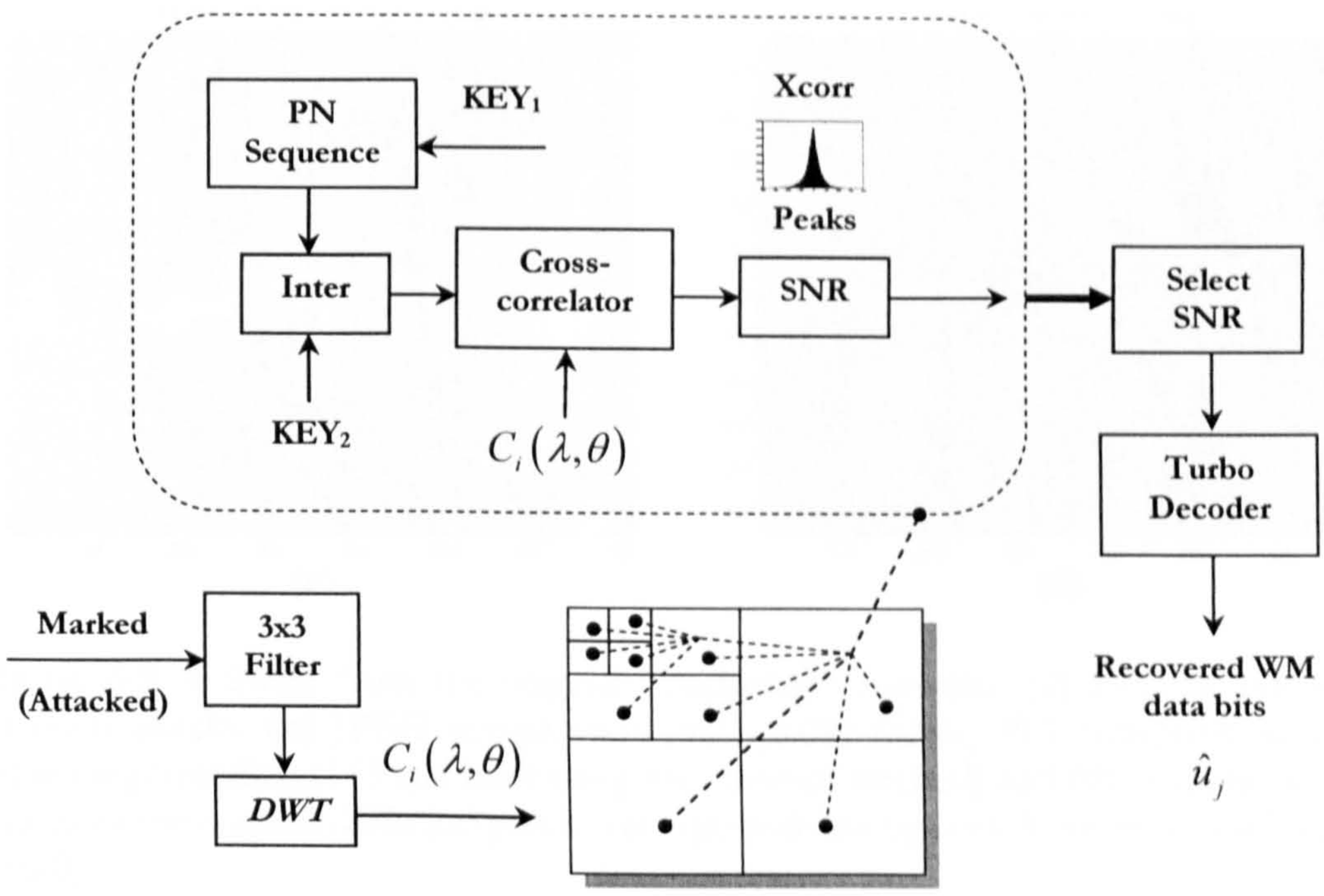


Figure 6-7 Wavelet-based watermark recovery.

each sub-band as an indicator of sub-channel quality. Different types of attack affect different levels and orientations in different ways, so it is always possible to select an optimal sub-band via SNR. In this way we can take advantage of any strong structure associated with the original image/video.

Correlation is therefore performed separately for each sub-band, obtaining a set of cross-correlation peaks (one peak for each embedded data bit) for each sub-band. A SNR is then computed for each set of cross-correlation peaks (section 4.3), and retrieval is carried out (only) for the sub-band with the highest SNR.

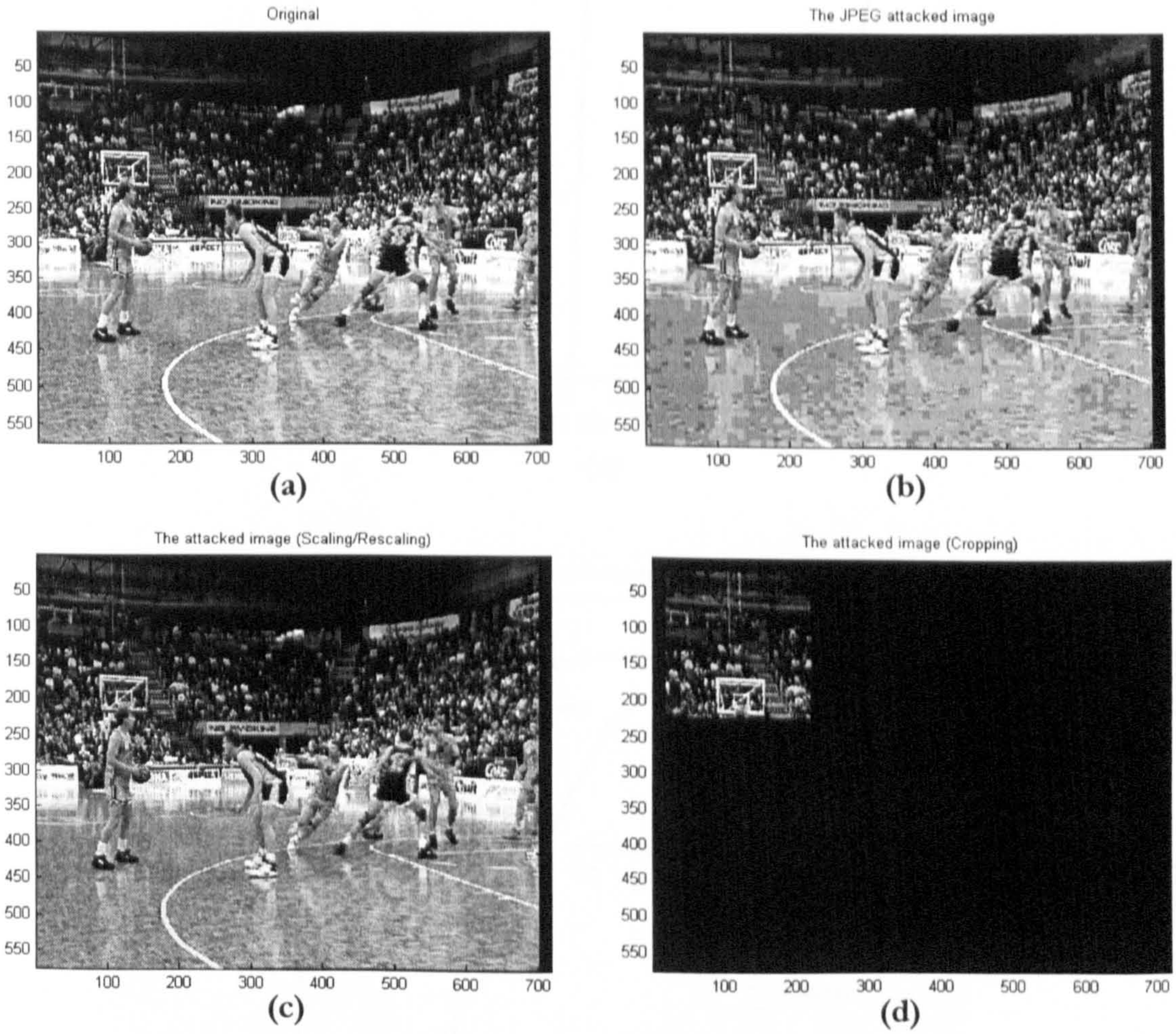
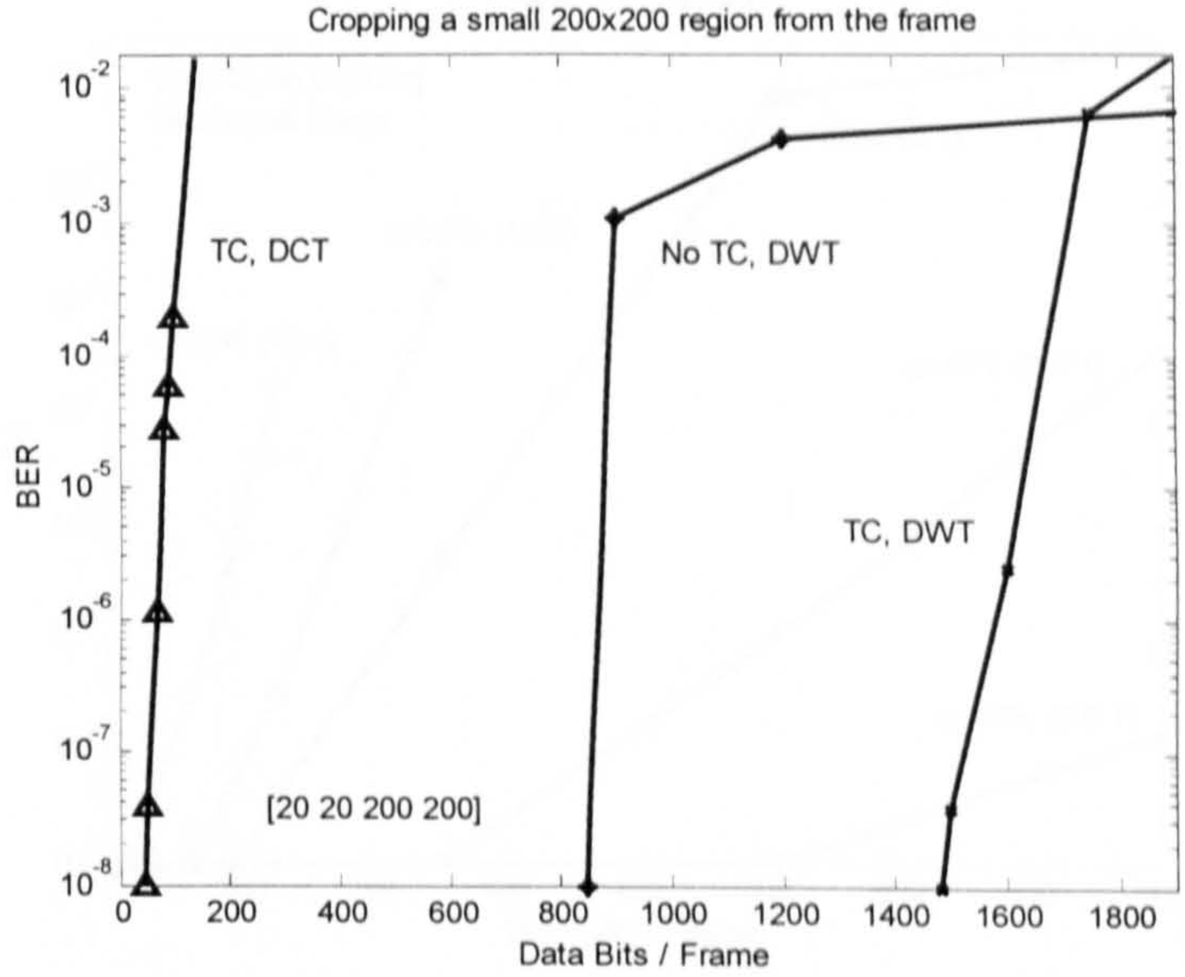


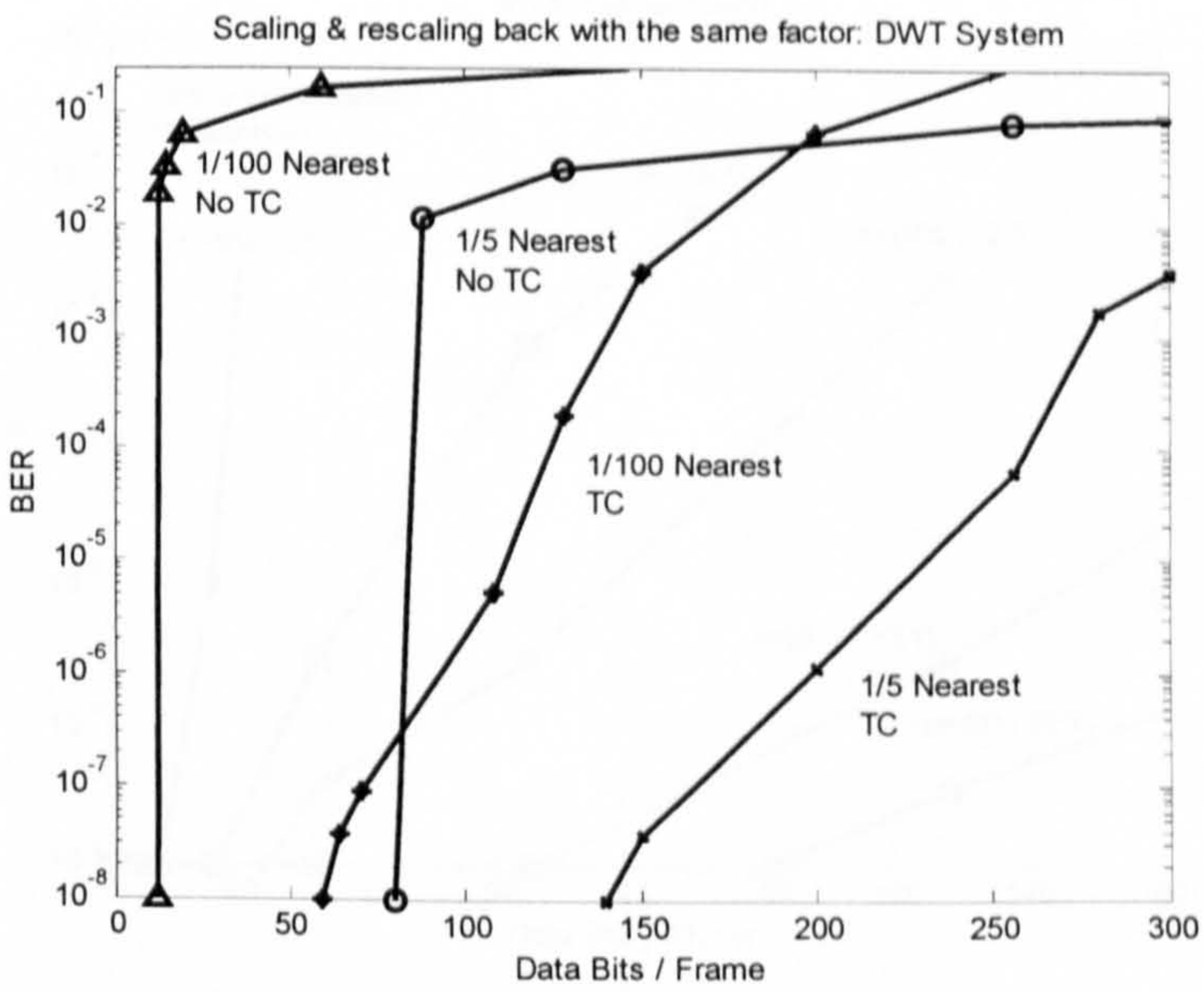
Figure 6-8 A frame from the original "Basketball" sequence **(a)** and the effects of different attacks: **(b)** JPEG compression (5% quality factor, 30:1 compression ratio), **(c)** scaling/rescaling (1/5 and back using the 'nearest' method) and **(d)** cropping a small area from the original (200x200 pixels rectangle with the upper left corner at the location [20,20]).

6.5 Performance of the Wavelet-based system

The visual effect of attacks like cropping, scaling/rescaling and JPEG compression is illustrated in **Figure 6-8**. The magnitude of these attacks is quite extreme, leading to unacceptable visual artefacts. The scaling for example was performed with a very bad quality



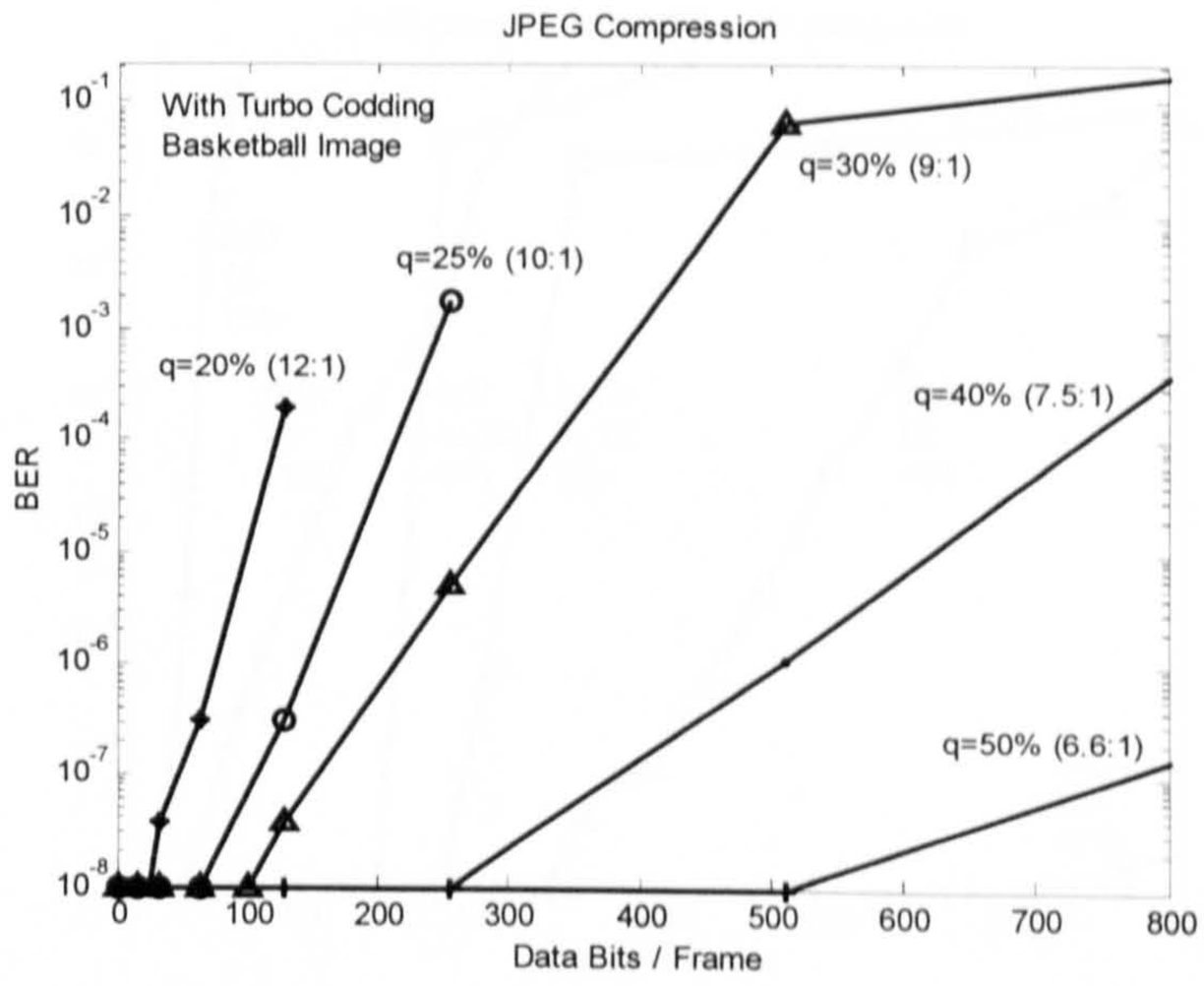
(a)



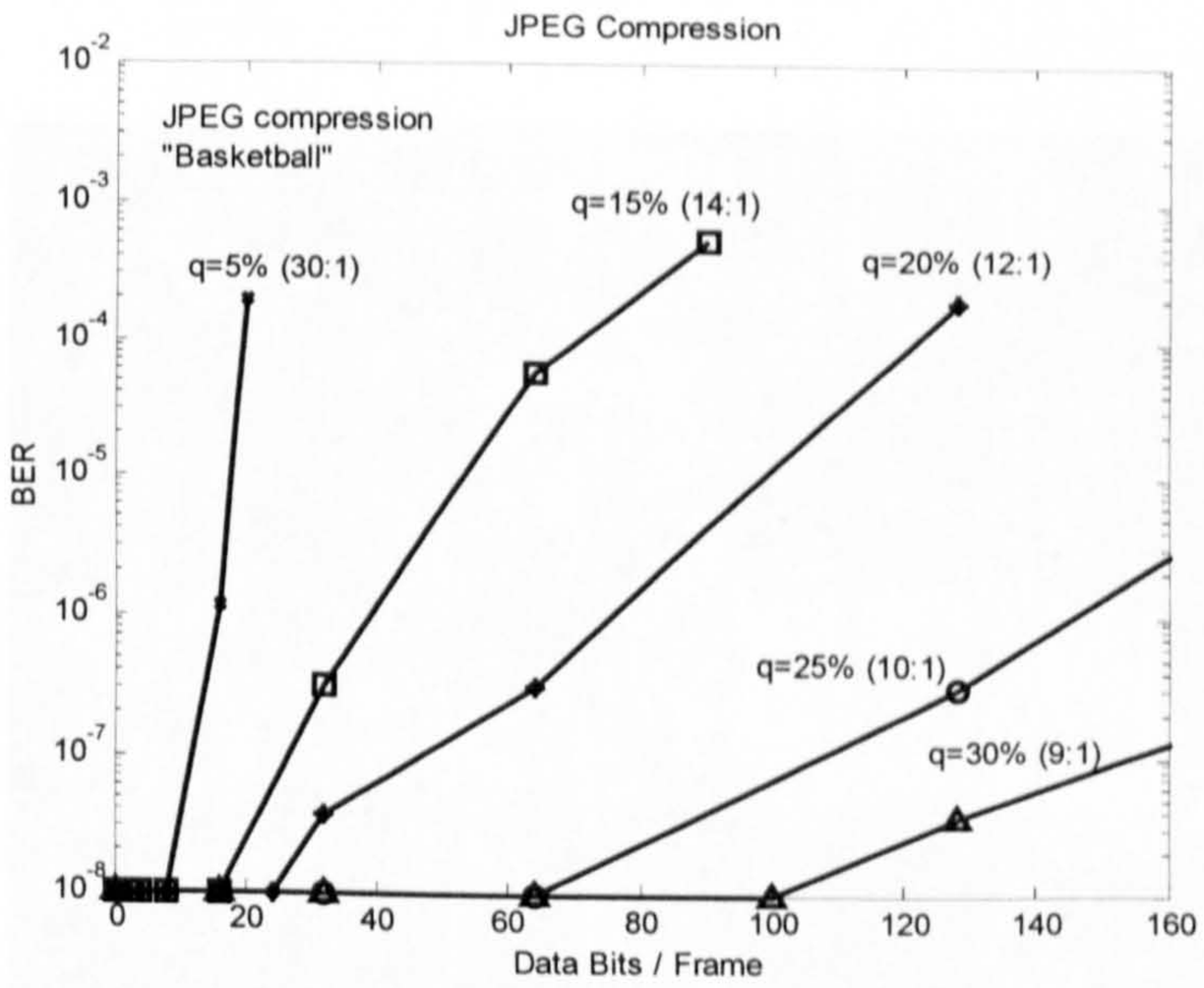
(b)

Figure 6-9 The performance of the DWT system for: **(a)** cropping and **(b)** scaling-rescaling.

interpolation filter, just to see how well the system performs. For JPEG compression, important artefacts become visible for a quality factor lower than about 25% (10:1 compression). **Figure 6-9 ↔ Figure 6-14** presents the performance of the DWT system for various attacks and some comparisons between the DCT scheme and the DWT scheme in order to outline the superiority of the DWT-based approach. The DCT scheme used for comparison is the one described in **Chapter 5**.



(a)



(b)

Figure 6-10 The performance of the DWT system for: **(a)** medium quality JPEG compression and **(b)** low quality JPEG compression.

As might be expected from the compact support, the most significant advantage of wavelets occurs under cropping and scaling. For cropping, a rectangle of 200x200 pixels was selected from the upper left corner of the frame, as shown in **Figure 6-8(d)**. This location was selected since it has average detail. Clearly, cropping to this degree is an extreme case and is unlikely to occur in practice. It is apparent from **Figure 6-9(a)** that the DCT scheme has poor performance even with FEC (TC, DCT curve), whereas the DWT scheme performs very well

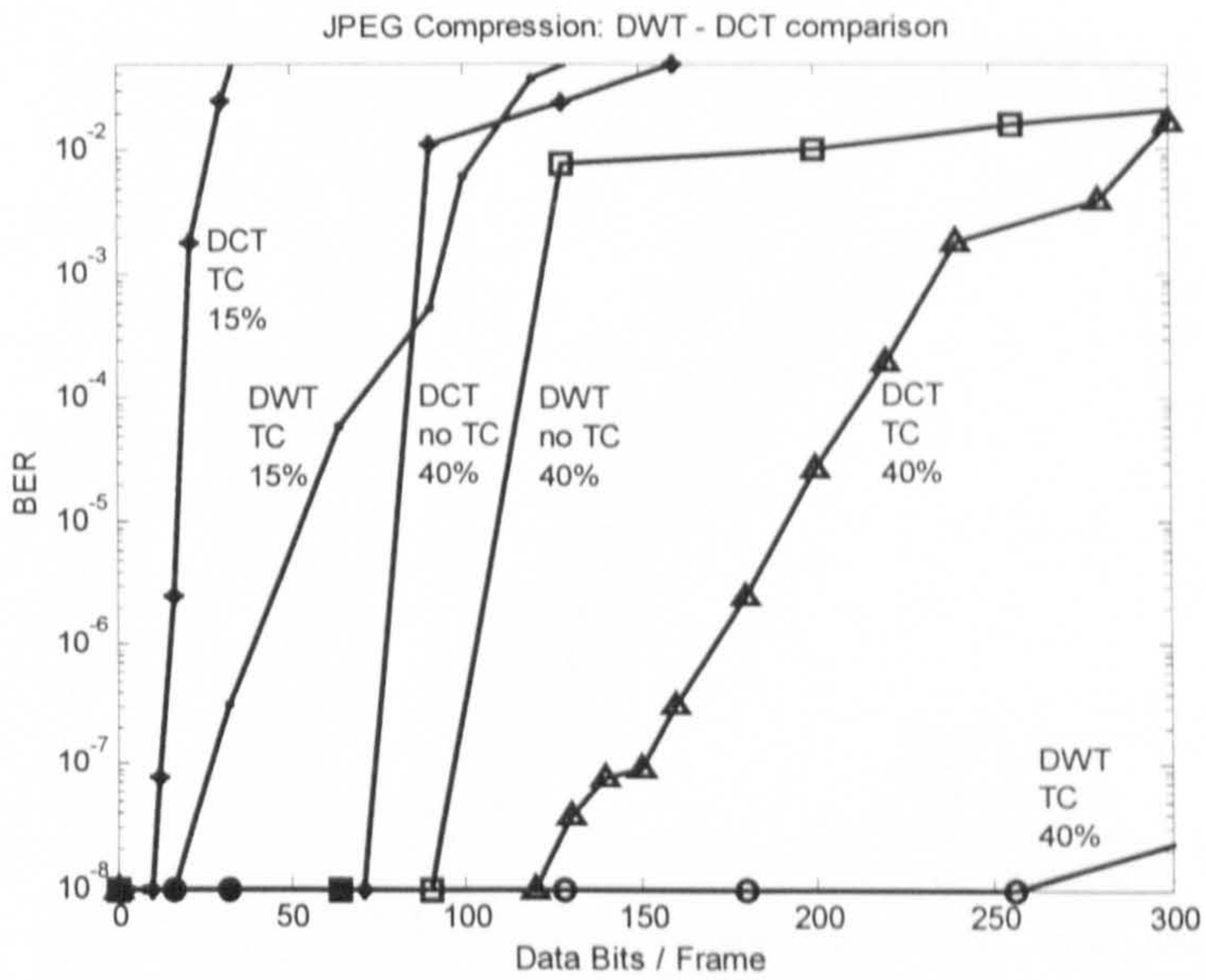


Figure 6-11 Comparison between the DCT and DWT systems for medium quality JPEG attack.

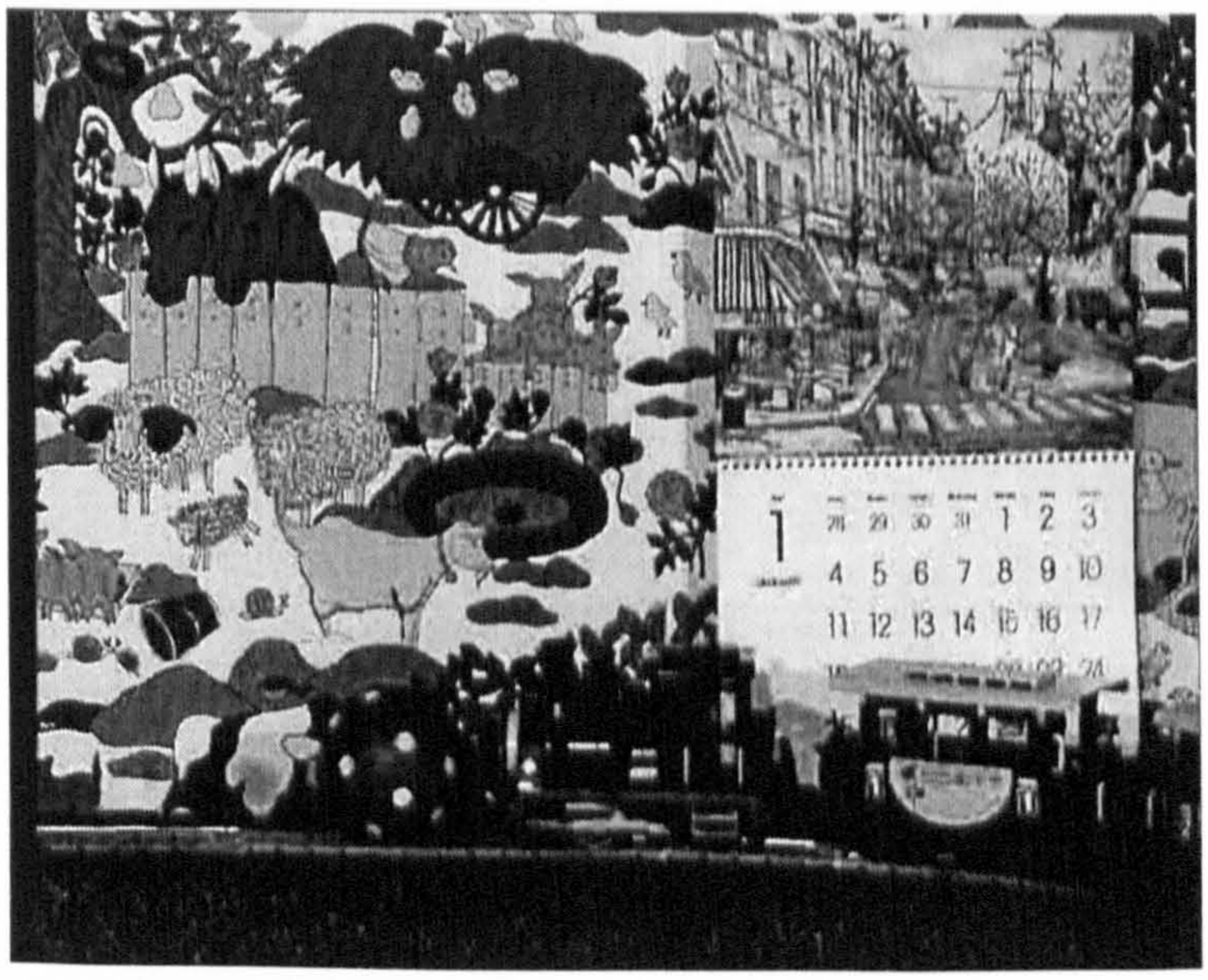
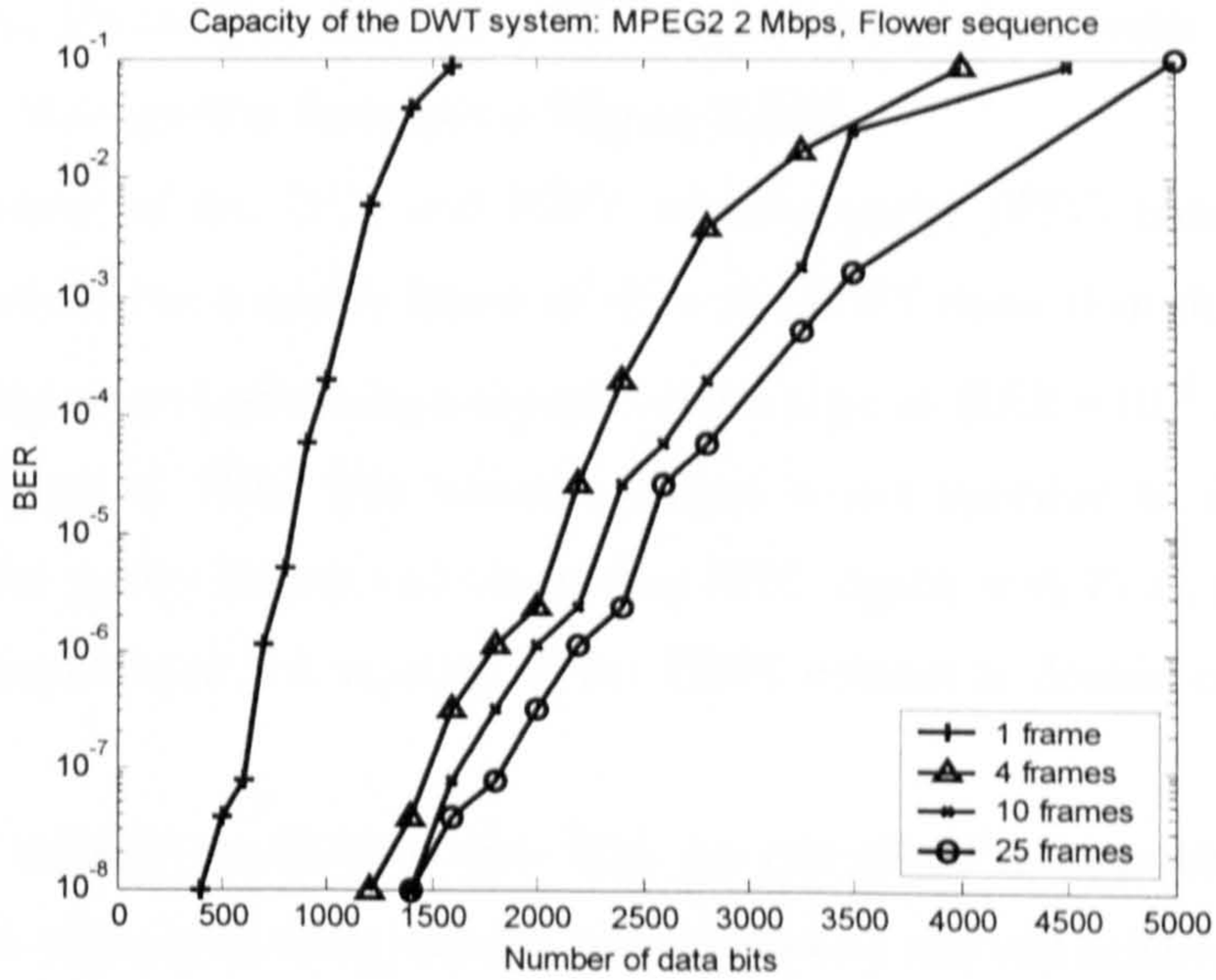


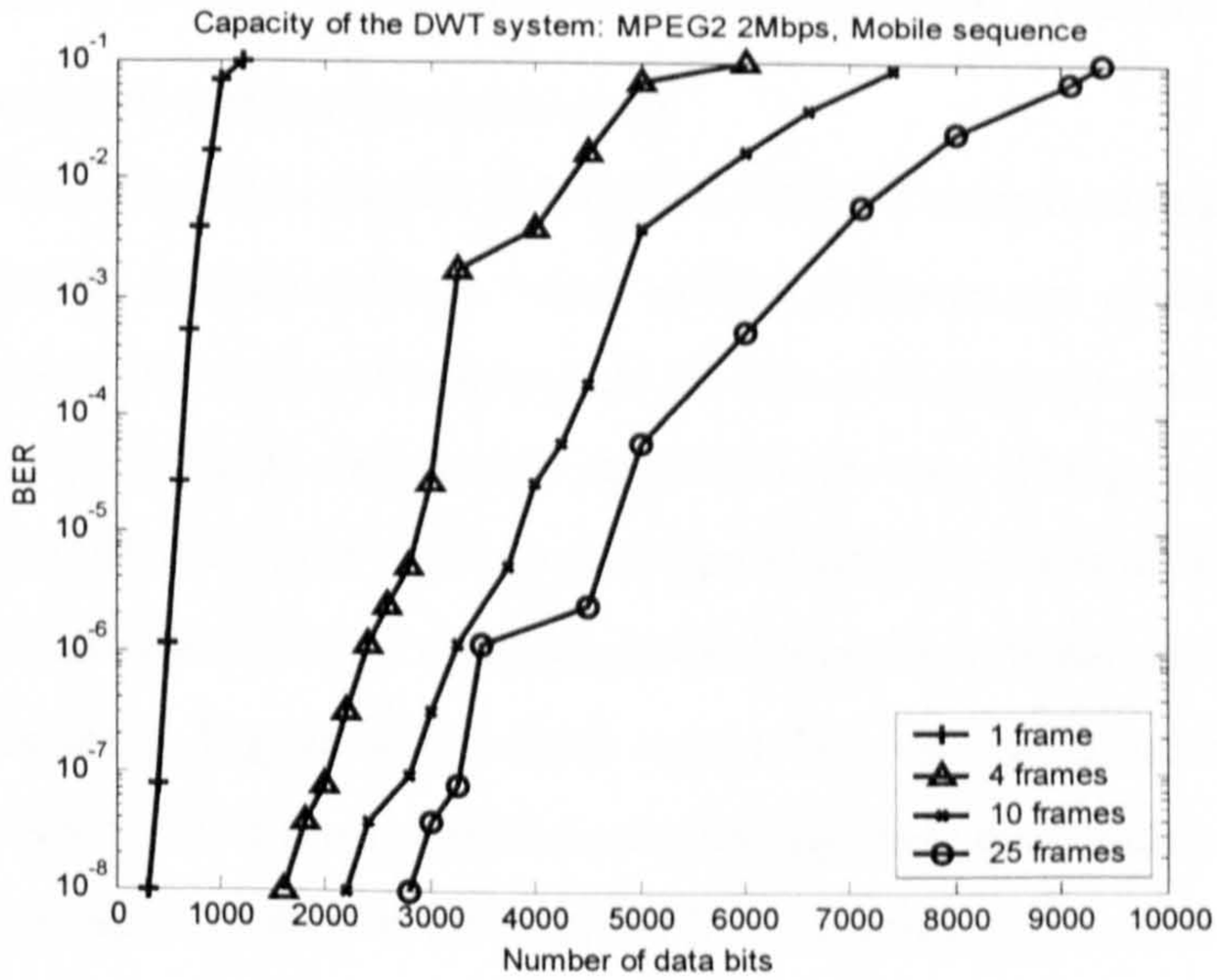
Figure 6-12 The "mobile" sequence MPEG2 compressed to 2Mbps: it is easy to spot the blocking artefacts even on a still frame.

without FEC (No TC, DWT curve) leading to over 20 kbps at $BER = 10^{-8}$. With FEC the capacity increases to 37 kbps, but will reduce markedly under a combined attack.

Figure 6-9(b) shows the results for scaling. The frame is scaled up or down and then brought back to the original size (720x576). Even so, with the worst kind of scaling, the DWT system performs quite well. The effect of this kind of attack results in luminosity changes and geometric distortion, **Figure 6-8(c)**. A DCT system can't cope with this attack. In contrast, the



(a)



(b)

Figure 6-13 Performance of the DWT system under 2Mbps MPEG2 compression attack for: **(a)** "flower" video sequence and **(b)** "mobile" video sequence.

DWT gives very acceptable performance, especially when using FEC. For example, for 1/5 "nearest" scaling, the capacity is about 80 bpf (bits per frame) which translates to 2 kbps (25 frames per second) and increases to about 140 bpf (3.5 kbps) with FEC.

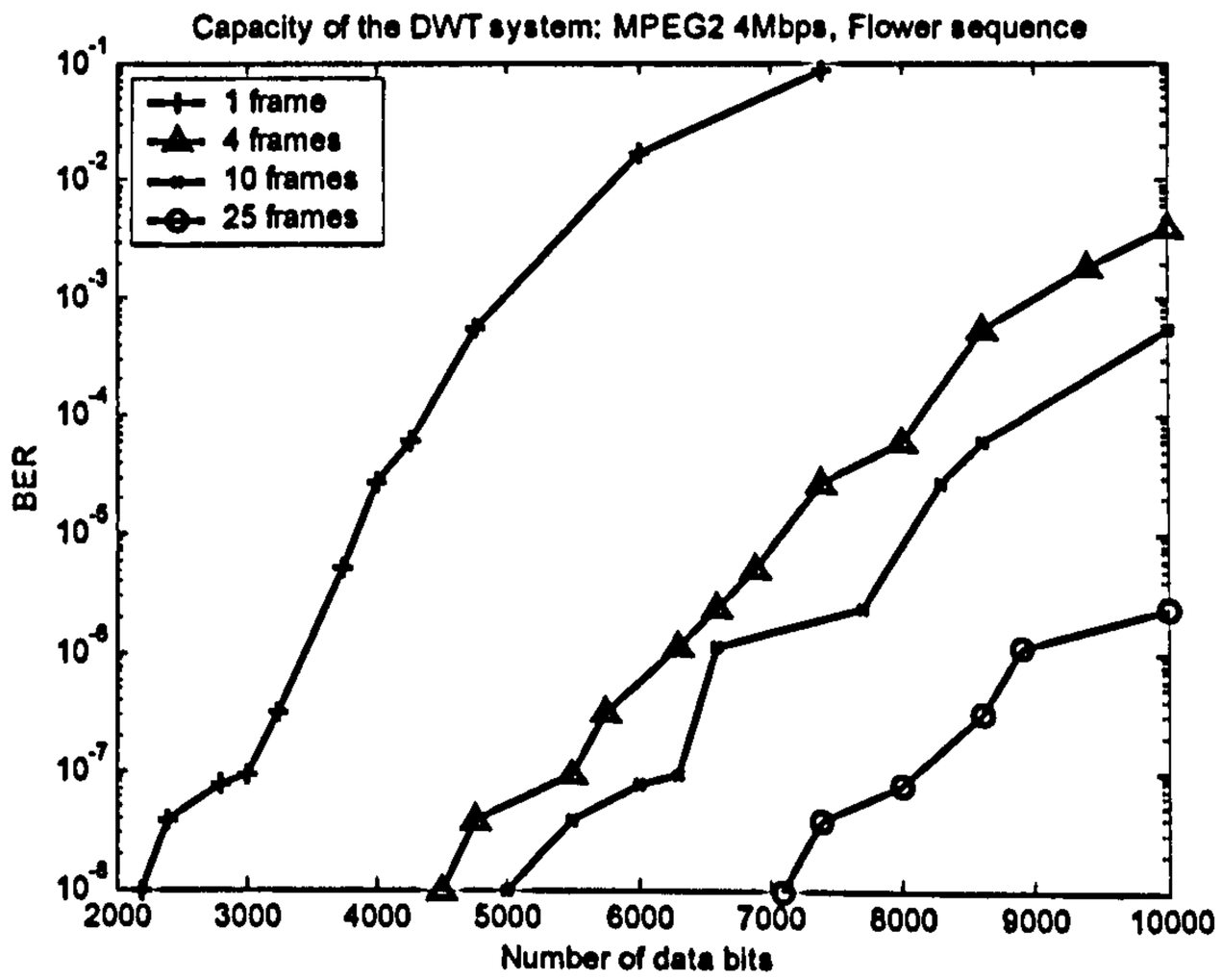
The results for JPEG compression with several different quality factors are presented in **Figure 6-10(a)** and **Figure 6-10(b)**. As **Figure 6-10(b)** indicates, for a relatively high compression factor of 10:1 (25% quality, slight visual artefacts) and with Turbo coding, the wavelet scheme can achieve a capacity of 64 bpf. Even under extreme JPEG compression (30:1 compression, 5% quality, with heavy blocking artefacts) the wavelet scheme still has a capacity of 8 bpf. This attack is illustrated in **Figure 6-8(b)**.

A comparison of the DCT and DWT schemes under JPEG compression attack is shown in **Figure 6-11**. For a quality factor of 40%, the DWT more than doubles the capacity when Turbo coding is used, achieving a capacity over 6 kbps at $BER = 10^{-8}$. This result clearly shows the advantage of FEC. The wavelet scheme is net superior to the DCT scheme, especially for higher quality factors and when using FEC. Again, with FEC, for a quality factor of 40% (7.5:1 compression) the capacity of the DWT scheme is double compared with the DCT scheme.

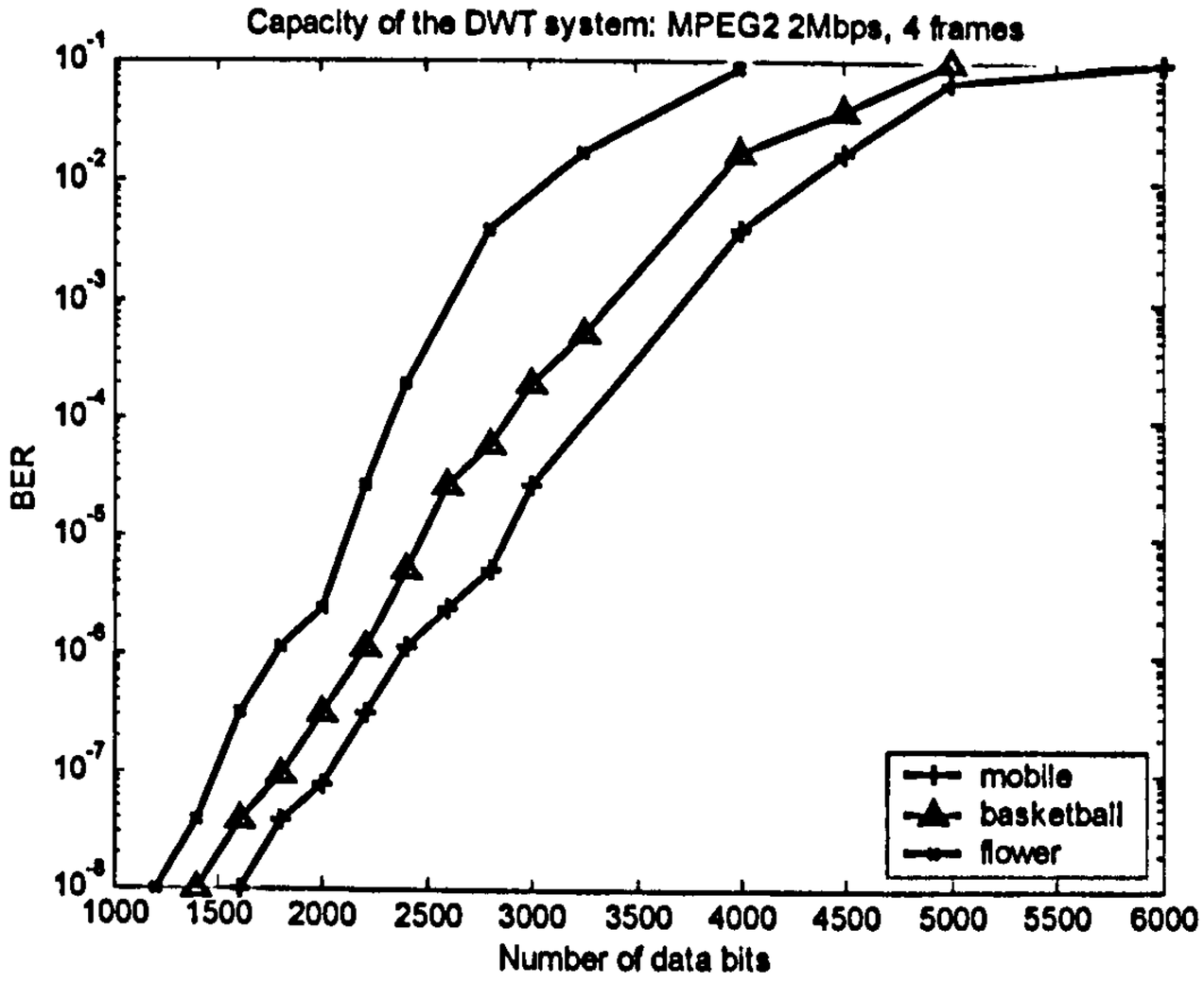
Since the capacity per frame is quite high, we can afford to increase the robustness in the expense of the capacity (trading capacity for robustness) and still achieve a higher capacity compared with the single frame case, by inserting the same watermark in a number of $n\,(n \le 25)$ successive frames. In this way the recovery is much simplified since takes place only once, and is easier to combat frame dropping.

This case is illustrated in **Figure 6-13(a)** for MPEG2 compression attack, which gives an impressive capacity of about 1Kbps, when at least 4 frames are averaged together. The improvement between the 4, 10 and respectively 25 frames averaging seems to be quite small, however this is due to the high compression applied in this case (2Mbps); for a medium level of compression the difference between these cases are much more obvious **Figure 6-14(a)**.

It must be said that MPEG2 compression at 2Mbps is a drastic attack, which leads to important visual artefacts **Figure 6-12**, which can be best seen in a moving sequence. The results mentioned above are quite remarkable considering that were obtained for the "Flower garden" sequence, which is notorious for its difficulty and usually gives worse results compared with the other test sequences. This can be seen very well comparing the results obtained for the "Flower garden" sequence with those of the "Mobile" sequence presented in **Figure 6-13(b)**. A direct comparison between three typical sequences is provided in **Figure 6-14(b)**.



(a)



(b)

Figure 6-14 Performance of the DWT system under MPEG2 attack for: **(a)** "flower garden" video sequence, compressed at 4Mbps and **(b)** different video sequences, compressed at 2Mbps, 4 frames averaging.

6.6 Conclusions

The use of the wavelet transform in digital watermarking has many advantages compared with the traditional FFT/DCT transform, fact very well illustrated by the performance of the DWT-based scheme.

The properties of the wavelet transform itself lead to a significant increase in capacity compared with the DCT based systems, in spite of the much simpler HVS used in wavelet's case. From the robustness perspective, wavelets also offer much better results. The results suggest that the DWT has significant advantages under attacks which are likely to be encountered in studios: compression, scaling and cropping.

Under JPEG compression attack, the DWT can more than double the capacity of a DCT system. Subjected to a MPEG2 compression attack, the DWT system can achieve capacities four times as much as the DCT system. For a typical scaling/re-scaling attack, a Turbo coded DWT scheme can yield capacities in excess of 1 kbps, whilst under the same conditions a DCT scheme fails. The DWT scheme has been found to be particularly robust to cropping: for example, the Turbo coded DWT scheme had a capacity of some 37 kbps, compared to 1 kbps for the DCT scheme.

The improved robustness of the DWT scheme is mainly attributed to the spatially local and spatially global support of wavelets. For example, wavelets with local support are less likely to be affected by cropping, compared to the theoretically infinitely long basis functions used in Fourier analysis. The multiresolution feature can also be exploited to optimize retrieval, by embedding all data bits in each sub-band and measuring sub-band SNR. The hierarchical, multiresolution nature of the DWT has also a fundamental advantage for embedding, by performing an analysis similar to that of the HVS. Since the DWT is a HVS model by its nature, even a relatively simple HVS model may suffice. Finally, the DWT has a computational advantage compared with the DCT and it does not suffer from the blocking artefacts so common to the DCT.

Chapter

7

"We will either find a way, or make one."

Hannibal, 247-182 BC

"It is not the answer that enlightens, but the question"

Eugen Ionescu, 1909-1994

Adding Robustness to Geometrical Attacks

One of the most difficult problems in digital video watermarking is watermark recovery in the presence of geometric attacks like frame shift, cropping, scaling, rotation, and change of aspect ratio, especially when some of these are combined together.

For uncompressed video, the geometric attacks tend to be less severe compared to those for image watermarking, mostly due to visibility considerations and due to the TV studio particularities [Cheveau et al, 2000]. On the other hand, the recovery problem is compounded for video since it must be carried out blind due to the difficulty of storing the original.

In this case, in order to re-establish synchronisation one could use sliding window cross-correlators, as described in Chapter 3 and Chapter 5. Unfortunately the search space grows very quickly, making it difficult to recover the watermark in a reasonable time (section 3.2.2 and section 5.4.2). Clearly, given that (ideally) retrieval in a video context must be done in near real time, the computational problem is very significant in the presence of attacks.

This chapter provides a solution[1] to this difficult problem by employing an additional reference watermark and image registration techniques, while maintaining in the same time the high capacity and the robustness of the Wavelet based system presented in Chapter 6.

[1] The system can cope with all geometric attacks specified in the EBU recommendations [Cheveau et al, 2000], with the exception of bending/shearing attacks which are still an open problem. In the actual configuration, this problem can be solved by employing a sliding window correlator.

7.1 Methods of Combating Geometrical Attacks

Invariant transforms

One of the answers of the research community to this difficult problem was to use transforms invariant to these attacks.

For example, embedding a watermark in the magnitude of the DFT coefficients yields a shift invariant system, therefore eliminating the need of a 2-D sliding correlator, which normally would be used to recover this type of attack. Unfortunately, as section 2.4.3 already mentioned, due to its disadvantages, a FFT-based watermarking scheme is not the best choice for digital watermarking.

The Fourier-Mellin transform (FMT) was first used in [O'Ruanaidh et al, 1998] to achieve rotation, scaling and translation invariance for image watermarking. The FMT doesn't have a fast direct computing algorithm, but can be simulated by transforming the Cartesian coordinates into log-polar coordinates and then performing a Fourier transform of this output.

Unfortunately, marking in the FMT domain has two major drawbacks: the need to compute the inverse log-polar transform which is a lossy operation that drastically reduces system performance, and the need to maintain the FFT symmetry, which halves the watermark capacity. Due to these disadvantages the technique offers only poor results, as even the author acknowledged later on. An improved technique was later proposed in [Lin et al, 2000].

A very promising technique is suggested in [Loo et al, 2000], which uses the Complex Wavelet Transform (CWT) domain for watermarking. The dual-tree CWT transform offers shift invariance, directionally selective filters and limited redundancy [Kingsbury, 1998 and 1999]. One of the most appealing features of the CWT is that the phase of the CWT coefficients can be used to infer pixel shifts quite accurately. Therefore using a registration algorithm based on motion estimation, one could combat even the non affine geometric distortions produced by StirMark [Loo et al, 2000].

Special watermark arrangements

Another possibility is to cleverly arrange the watermark so that the search space reduces considerably. Some examples could be: circularly symmetric watermarks [Solachidis et al, 1999], tiled (cyclic) watermarks [Kalker et al, 1999-1 and 1999-2] and [op de Beeck et al, 2001], self-similar watermarks [Dittmann et al, 2000]. This approach is in general more successful than the previous method, but still has a major drawback: although the geometric

attacks are more easily handled, the capacity of such a system is relatively limited, due to the special arrangement of the watermark.

Reference watermarks

A slightly different approach is to use reference watermarks. These are separate watermarks additional to the main watermark which carries the actual watermark data. The reference watermark is in fact not carrying any information at all, being used only as a countermeasure for geometric attacks. They are sometimes called templates or patterns. In this scenario, the image/video sequence will carry two distinct watermarks, on top of each other, and which are normally embedded in the same domain in order to keep the complexity of the method low. The reference watermarks can be used to identify the parameters of the (affine) geometrical attack. Once these parameters are identified it is possible to undo the geometrical attack and then recover the main watermark.

Most of the existing schemes are using a combination between these two later methods [Kalker et al, 1999-1 and 1999-2], [op de Beeck et al, 2001], [Pereira et al, 1999]. Usually the reference is embedded in a block wise manner (the size of the template is smaller than the size of the image/frame) [Kalker et al, 1999-1 and 1999-2], [Pereira et al, 1999] in different locations of the image/frame, so the peaks obtained after the cross-correlation will be distributed on a grid. Their position on the grid depends on the attack, and therefore it is possible to estimate (identify) the parameters of the affine geometrical attack using this grid [Kalker et al, 1999-1 and 1999-2], [Wolberg et al, 2000].

In order to get shift invariance, the reference (and the main watermark as well) is sometimes embedded in the Fourier domain [Pereira et al, 1999]. One could achieve robustness to scaling and rotation attacks by using log-polar transforms which translate the scaling and rotation to spatial shifts which can then be easily recovered. Robustness to aspect ratio changes can be achieved by using log-log transforms to convert the scale changes to spatial shifts. Compared with watermarking in the Fourier–Mellin domain, this approach has the advantage that doesn't require the computation of the lossy inverse Fourier-Mellin transform. In fact this technique is very close to the well known image registration and template matching problem.

[Pereira et al, 1999] partially uses this method, but instead of following the normal template matching approach they transform it in a point-matching problem over a log-polar or log-log map which involves a limited exhaustive search. The both watermarks are embedded in

the Fourier domain and therefore the scheme inherits the disadvantages of Fourier based watermarking techniques described in section 2.4.3.

Moreover embedding two watermarks in the same domain raises two problems: the possible interference between the watermarks and the efficiency of the embedding. In order to counteract the first problem, the watermarks have to be orthogonal, and therefore the PN sequence corresponding to each watermark has to be carefully selected. The second problem arises because in this case, two watermarks have to "share" the maximum value given by the visual model for modifying a certain coefficient. There can be only a limited amount of modification allowed for a certain coefficient, in order to maintain the invisibility of the watermark(s). So there is always a problem in finding the right balance between the strength of each watermark, and therefore the energy of each watermark is lower than when embedding only a single watermark.

7.2 Symmetrical Phase-Only Matched Filtering

Another way of using the advantages of the Fourier transform is to use it for implementing fast cross-correlators (section 2.4.3). The roots of FFT cross-correlation can be found in optics, where a lens basically performs a Fourier transform. Then the concept was then quickly adopted by the image processing community.

Cross-correlation or matched filtering is particularly used for pattern (template) matching, image registration and recognition and motion estimation. The aim of matching is either to determine the presence of a template/image in a noisy scene (pattern recognition), or to determine the parameters of a geometric transformation relating two images (image registration). Such a FFT based cross-correlator can easily recover 2-D shifts, saving allot of computing time compared with the classical cross-correlators [Kuglin et al, 1975], [Horner et al, 1984], [Pech-Pacheco et al, 199x], [Chen et al, 1994] and [Hill et al, 1999].

Let's consider two images $f_1(x,y)$ and $f_2(x,y)$ which differ only by a displacement (x_0,y_0). This can be expressed as $f_2(x,y)=f_1(x-x_0,y-y_0)$. The Fourier transforms of these images are $F_1(u,v)$ and respectively $F_2(u,v)$. It is well known that the output of a classical matched filter is primarily dependent on the energy of the image rather than its own spatial structures. This is why the matched filter provides a relatively poor discrimination between objects of different shapes but similar size or energy content. Furthermore, the filter

output is proportional with the image auto-correlation and the shape of the filter output around its maximum (x_0, y_0) is broad. Therefore locating this maximum in the presence of noise is relatively difficult.

The answer to this problem is the *Phase Only Matched Filter* (POMF), who's transfer function is equal with the spectral phase of the image [Chen et al, 1994]

$$H_{POMF}(u,v) = Phase\left(F_1^*(u,v)\right) = \exp\left(-j\phi_1(u,v)\right) \tag{7.1}$$

where $j^2 = -1$ and $\phi_1(u,v)$ is the spectral phase of the image $f_1(x,y)$. Since the spectral phase preserves the location of the objects but is insensitive to the image energy, the application of a POMF to a pair of identical images under the constraint of translation will result in a much sharper peak than in the case of the classical matched filter. This is very well illustrated in [Horner et al, 1984]. This explains the popularity of the POMF in the image processing community: the detection and location of the maximum is easier and therefore the POMF allows much better discrimination between different objects.

A further improvement of the POMF can be achieved by extracting and correlating the phases of both input images. In this case the filter can be defined as

$$H_{SPOMF}(u,v) = \frac{F_2(u,v)}{\left|F_2(u,v)\right|}\frac{F_1^*(u,v)}{\left|F_1^*(u,v)\right|} = \exp\left[j\left(\phi_2(u,v) - \phi_1(u,v)\right)\right] \tag{7.2}$$

where the $\phi_1(u,v)$ and $\phi_2(u,v)$ are the spectral phases of the image $f_1(x,y)$ and respectively $f_2(x,y)$. In the absence of noise, this reduces to

$$H_{SPOMF}(u,v) = \exp\left[-j2\pi\left(ux_0 + vy_0\right)\right] \tag{7.3}$$

The output of the filter is given by the inverse Fourier transform of $H_{SPOMF}(u,v)$ which is in fact a Dirac δ function, centred at the location (x_0, y_0). This filter is called a *Symmetrical Phase Only Matched Filter* (SPOMF) and can be seen as a two step process: first the extraction of the phases of the input images and then the phase only matched filtering. An assessment of the comparative performance of different matched filters can be found in [Chen et al, 1994].

7.3 Image Registration and Watermarking

By using such a Phase Only Matched Filter, the shift problem is easily solved. Unfortunately, this technique can be applied only to frame shifts.

On the other hand, it is well known in image processing that transformation of Cartesian coordinates into log-polar coordinates converts scale and rotation to spatial shifts [Casasent et al, 1976-1 and 1976-2], [Chen et al, 1994], [Reddy et al, 1996], [Cheng et al, 1998], [Wolberg, et al, 2000]. As stated before, performing the Fourier transform of this output is in fact equivalent with performing the Fourier-Mellin transform (FMT) of the original image, and the advantage of the Fourier-Mellin transform is its scaling and rotation invariance.

The idea developed by Casasent and Psaltis [Casasent et al, 1976-1 and 1976-2] was quickly adopted for image processing in the context of *image registration*. This involves two images; the original and an attacked copy, and the objective of an image registration module is to determine the parameters of the geometric distortion. The attack can then be inverted to give geometric alignment of two images.

When registering two images, typically the noise is relatively small, and so the correlator usually performs very well. However, the problem is more difficult for video watermarking since the original video frame is not available and "blind" recovery is necessary. In this case one can use spread spectrum watermarking to compensate for the unavailable original. Assuming that this (reference) watermark is a one bit watermark having the size of the (original) image, then is possible to see a correspondence with the classical image registration problem. The "original image" corresponds to the PN sequence used to embed the watermark, and the "attacked image" corresponds to the unsynchronised (attacked) watermarked image. This "attacked image", which is in fact nothing else than the watermarked image (possibly attacked) is composed by two components: the first component is the reference watermark and the second component is the image/video itself. The reference watermark represents the signal and the image/video itself can be regarded as (additive) noise. In the case of video watermarking/registration, the signal to noise ratio is therefore very low relative to that for typical image registration. By making a parallel with the watermarking, this technique can be called *"blind" registration*.

7.4 Log-Polar and Log-Log Mapping

As already mentioned, the highly desired geometric invariance can be achieved by using the FMT to convert rotation and scale to spatial shifts, which are then easily recovered with a SPOMF cross-correlator. The FMT itself does not have a fast algorithm, but as [Casasent et al, 1976-1 and 1976-2] shows, a simple variable change $x = \exp\xi$ on the input $f(x)$, followed

by a Fourier transform will yield the FMT. For the bi-dimensional case, this is equivalent to a log-log transform of the input, and permits recovery from arbitrary scale changes (aspect ratio changes).

Consider the case of arbitrary scaling with a factor *(a, b)*. If frame f_2 is the scaled replica of frame f_1 with a factor *(a, b)* then

$$f_2(x,y) = f_1(ax,by) \tag{7.4}$$

and its Fourier pair is

$$F_2(u,v) = \frac{1}{|ab|} F_1(u/a, v/b) \tag{7.5}$$

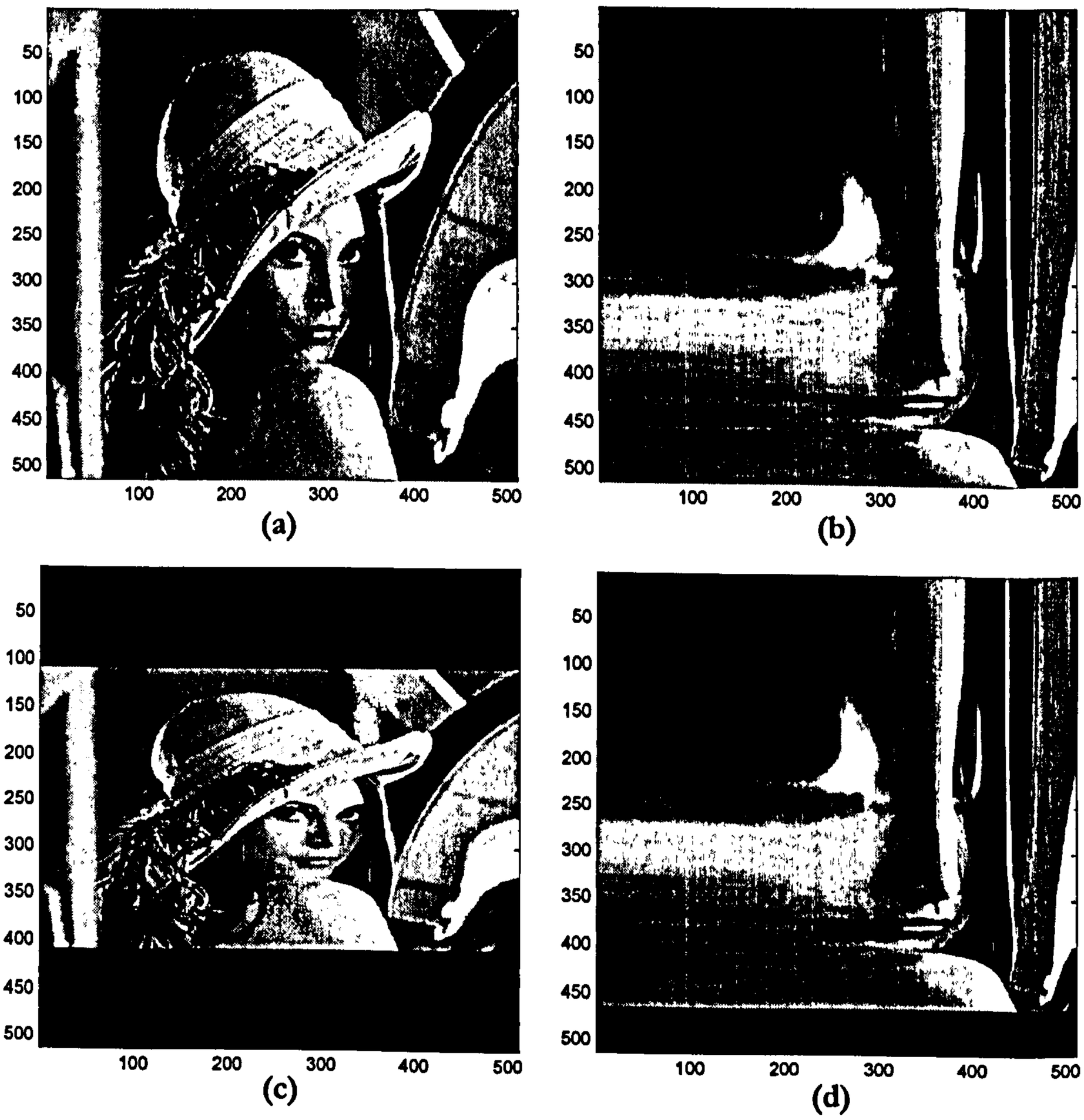


Figure 7-1 The log-log transformation and its results: **(a)** the original Lena image; **(b)** the log-log transformation of (a); **(c)** aspect ratio change attack and **(d)** the log-log transform of (c), the vertical scaling was transformed into a vertical shift in log-log coordinates.

If the Cartesian coordinates are transformed into *log-log* coordinates, and the multiplicative factor ignored, then

$$F_2(\log u, \log v) = F_1(\log u - \log a, \log v - \log b) \tag{7.6}$$

By using a SPOMF, the shifts $(-\log a, -\log b)$ can be found, and so the scale factors *(a, b)* are determined. An example illustrating this mechanism is provided in **Figure 7-1**. The scaling (aspect ratio change) is converted into spatial shifts. This can be easily observed by comparing the set of images **(a)** and **(b)** with their corresponding pair **(c)** and **(d)**.

Now consider f_2 to be a rotated replica of f_1, with angle θ_0

$$f_2(x, y) = f_1(x\cos\theta_0 + y\sin\theta_0, -x\sin\theta_0 + y\cos\theta_0) \tag{7.7}$$

$$F_2(u, v) = F_1(u\cos\theta_0 + v\sin\theta_0, -u\sin\theta_0 + v\cos\theta_0) \tag{7.8}$$

In order to convert this rotation into a shift, as in the previous case, the Cartesian coordinates are transformed into polar coordinates using

$$\begin{aligned} \rho &= \sqrt{x^2 + y^2} \\ \theta &= \tan^{-1}(y/x) \end{aligned} \tag{7.9}$$

This leads to

$$F_2(\rho, \theta) = F_1(\rho, \theta - \theta_0) \tag{7.10}$$

and the rotation can be easily recovered from the frequency domain shift.

Finally, when f_2 is both scaled with a factor *a* and rotated with angle θ_0, then

$$f_2(x, y) = f_1(a(x\cos\theta_0 + y\sin\theta_0), a(-x\sin\theta_0 + y\cos\theta_0)) \tag{7.11}$$

$$F_2(u, v) = \frac{1}{a^2} F_1((u\cos\theta_0 + v\sin\theta_0)/a, (-u\sin\theta_0 + v\cos\theta_0)/a) \tag{7.12}$$

In order to convert both scaling and rotation to shifts, it is necessary to convert the Cartesian coordinates into log-polar coordinates, using the following equations

$$\begin{aligned} x &= e^{\log\rho}\cos\theta \\ y &= e^{\log\rho}\sin\theta \end{aligned} \tag{7.13}$$

The result is

$$F_2(\log\rho, \theta) = F_1(\log\rho - \log a, \theta - \theta_0) \tag{7.14}$$

where the scale and rotation factors can be retrieved by SPOMF correlation. **Figure 7-2** shows the log-polar mapping and its effects. In this case the rotation is converted into a spatial shift. Again, this can be easily observed by comparing the set of images **(a)** and **(b)** with their corresponding pair **(c)** and respectively **(d)**.

Using the shift invariance property of the Fourier transform, it is possible to recover even combined attacks like shift combined with rotation and scaling or shift combined with aspect ratio changes. But since the FMT is not shift invariant, it is necessary to apply the Fourier magnitude of the frame (rather than the frame itself) to the input of the log-polar conversion module. The Fourier magnitude is shift invariant and so the rotation and scaling parameters can be found even in the presence of shift. After undoing rotation and scaling, the shift is then recovered by performing a simple SPOMF correlation. This technique works well in the particular case of image-image registration [Reddy et al, 1996], since the correlation peaks are relatively large and the phase loss can be tolerated. Unfortunately, for video watermarking

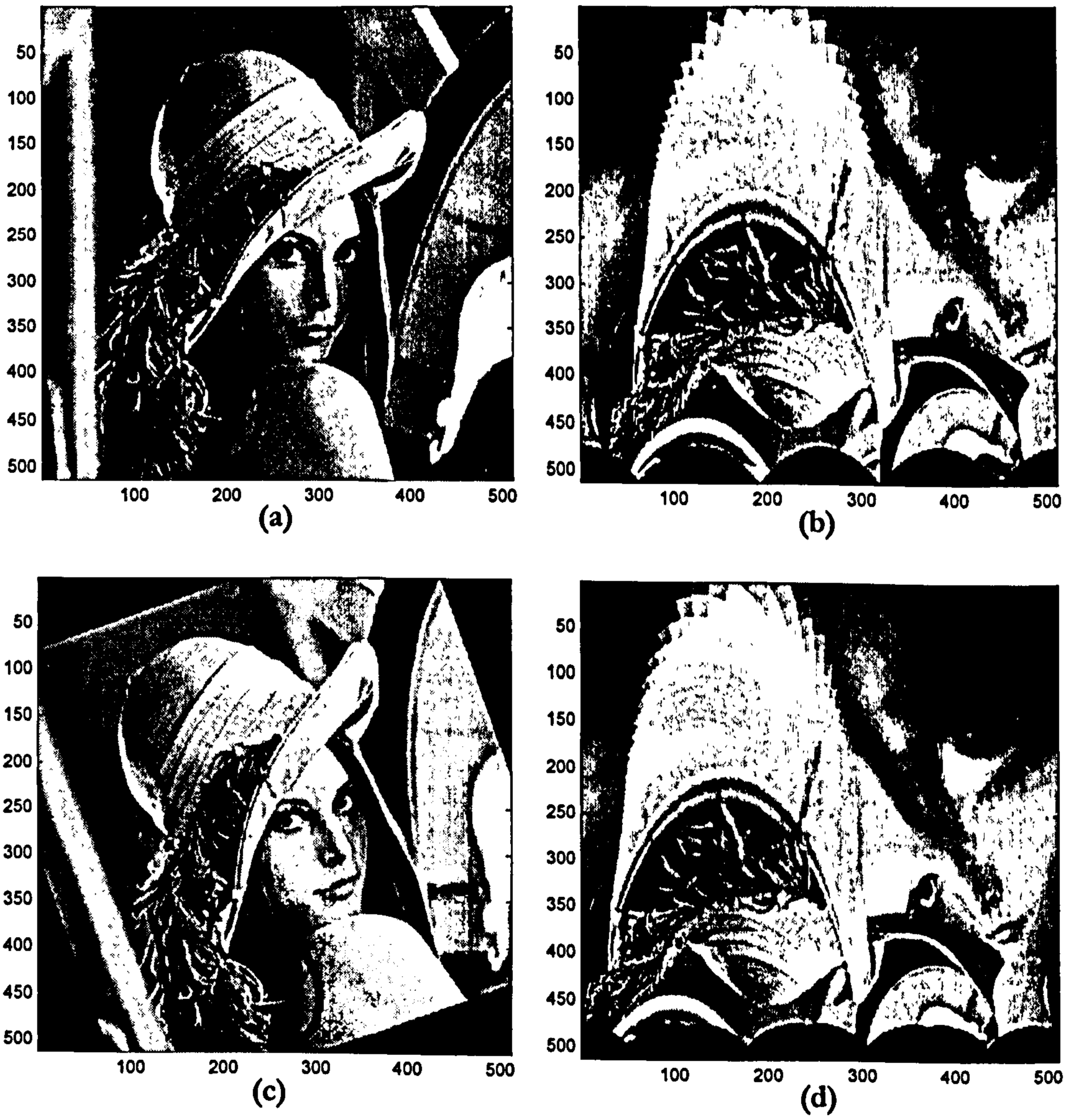


Figure 7-2 The log-polar transformation and its effects: **(a)** the original Lena image; **(b)** the log-polar representation of (a); **(c)** rotation attack and **(d)** the log-polar representation of (c), where the rotation was converted to a spatial (right) shift in the log-polar representation.

(blind registration), the loss can make cross-correlation unreliable, and this approach cannot be used for retrieval under combined attack.

A log-polar map permits recovery over a wide range of scale changes, rotation, or even combined scale-rotation attack. If a log-log map is used, then it is possible to recover arbitrary aspect ratio changes. The shifts alone are easily recovered using a SPOMF module. However, shift recovery from a combined attack (e.g. shift combined with scaling and rotation, or shift plus aspect-ratio change) requires a comprehensive search for all of the possible shifts [Wolberg et al, 2000], and is computationally intensive.

7.5 A High Capacity, Robust System

Using the method described in section 7.3, two different watermarks must be embedded. As section 7.1 mentioned, having two watermarks embedded in the same domain has some inconvenient. To overcome this problem, each watermark is embedded in a different domain. The first one, is a 1-bit watermark used exclusively for geometric reference, and for simplicity is embedded in the spatial domain. The second, multi-bit watermark is used for the data payload, and is embedded in the DWT domain. The advantage of using such an approach is obvious: the watermarks are orthogonal since they are embedded in different domains, so the cross-talk between them is minimal. Each watermark is embedded with the full strength dictated by its own visual model and overall the resulting system is relatively simple. Moreover the system combines the advantages offered by the blind video registration with the clear advantages offered by the Wavelet based embedding (Chapter 6). By combining these two

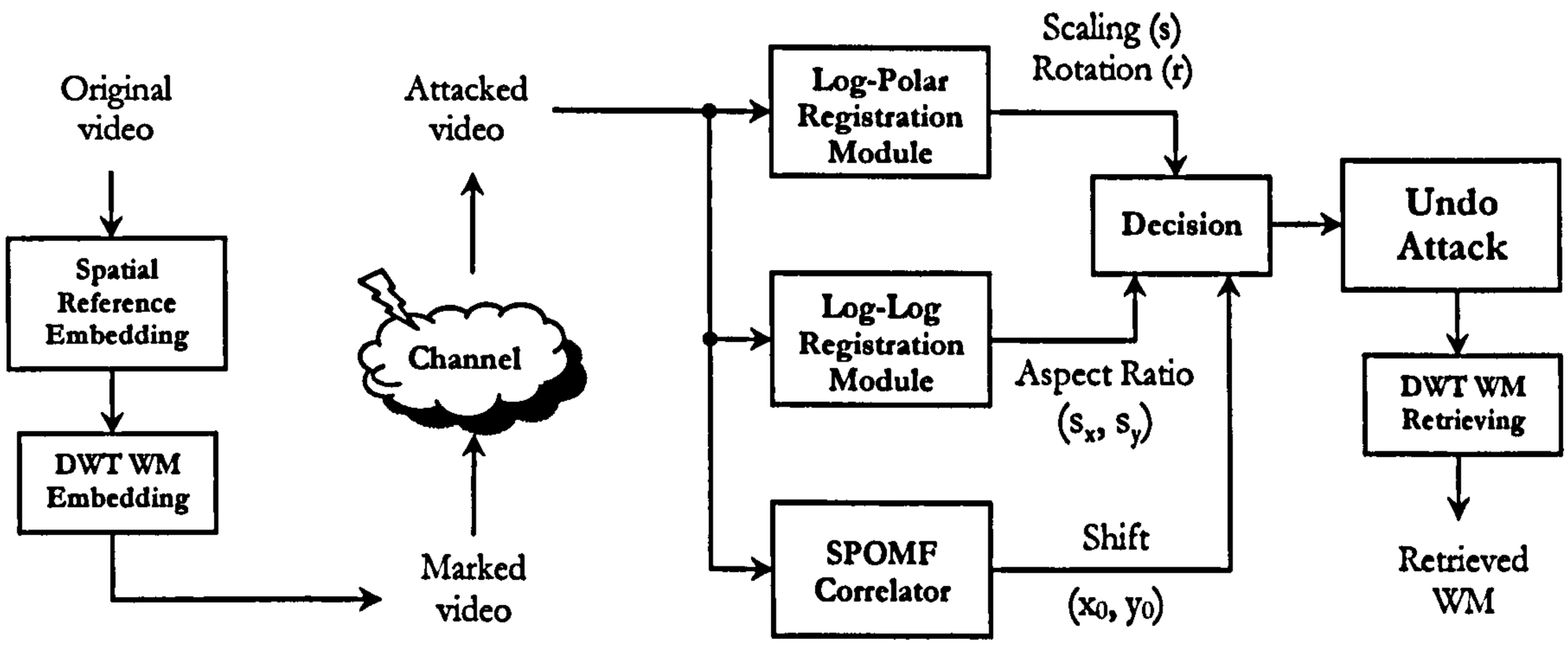


Figure 7-3 Block schematic of the geometric invariant video watermarking system.

techniques not only that the system ensures a good robustness against a wide range of attacks, but this can be achieved while maintaining the high capacity of the scheme intact.

The proposed system is presented in **Figure 7-3**. The first step is to embed the spatial reference watermark. Although the spatial domain is not the best place to cast a watermark, the reference is quite robust because effectively only one data bit is embedded in the entire frame. Since the capacity is not an issue in this case, the spatial domain can be successfully used and by doing so, one can exploit its biggest advantage: simplicity. This is embedded using the classical spread spectrum approach, according to a simple visual model that inserts a stronger watermark in those regions where it is less easily observed (at edges and in high texture regions). The same reference watermark is embedded in all the frames in order to increase the SNR at the correlator input via frame averaging. As a result, the registration takes place only once, and not for each separate frame. This is possible because attacks must be identical for each frame in order to avoid temporal artefacts.

The second step consists in embedding the high capacity watermark, according to the scheme described in Chapter 6.

At the retrieval side, the system employs three registration modules: the log-polar registration module which takes care of the rotation and scaling attacks, even when these two attacks are combined together; the log-log module which handles the aspect ratio change attacks and finally, a simple SPOMF correlator which handles attacks like spatial shifts and cropping, together with non-geometric attacks like compression or even a combination of these attacks.

The decision block determines if the reference watermark is present (to within a desired false detection probability), and if present it automatically determines the attack parameters. Once the parameters of the attack are identified, the attack is reverted back and the main watermark is then recovered.

Another advantage of using two watermarks is now apparent: if the reference cannot be found, one can assume that either the video is not marked, or that the mark is destroyed, and recovery of the main watermark payload can be abandoned (saving computation time).

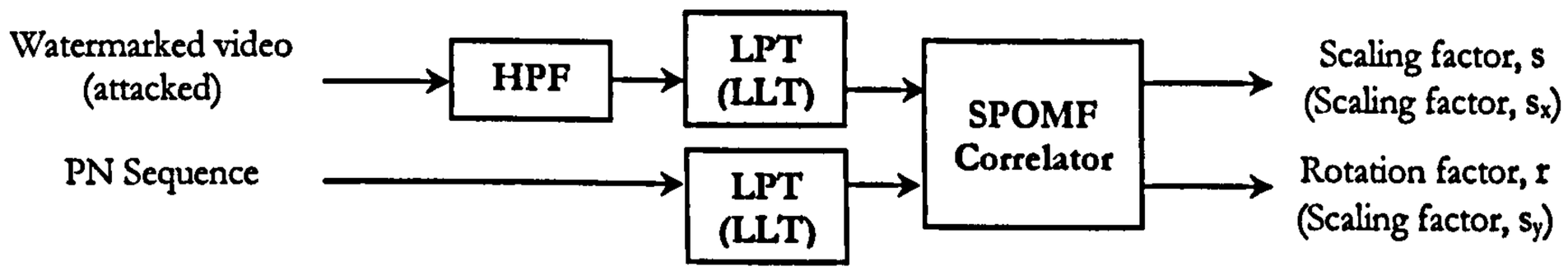


Figure 7- 4 The log-polar / log-log registration module.

Figure 7- 4 shows implementation detail of the log-polar transform (LPT) and the log-log transform (LLT) registration module. The role of the Laplacian high pass filter (HPF) is to remove low and medium frequency video components (which represent noise) and pass only the high frequency components, which contain the spread spectrum, noise-like watermark. This significantly improves the correlator performance.

7.6 Performance of the System

The registration module provides invariance to frame shift, rotation, scaling, rotation combined with scaling, and aspect ratio change. The system can also handle a range of other attacks, such as cropping, shift combined with cropping, compression, shift combined with

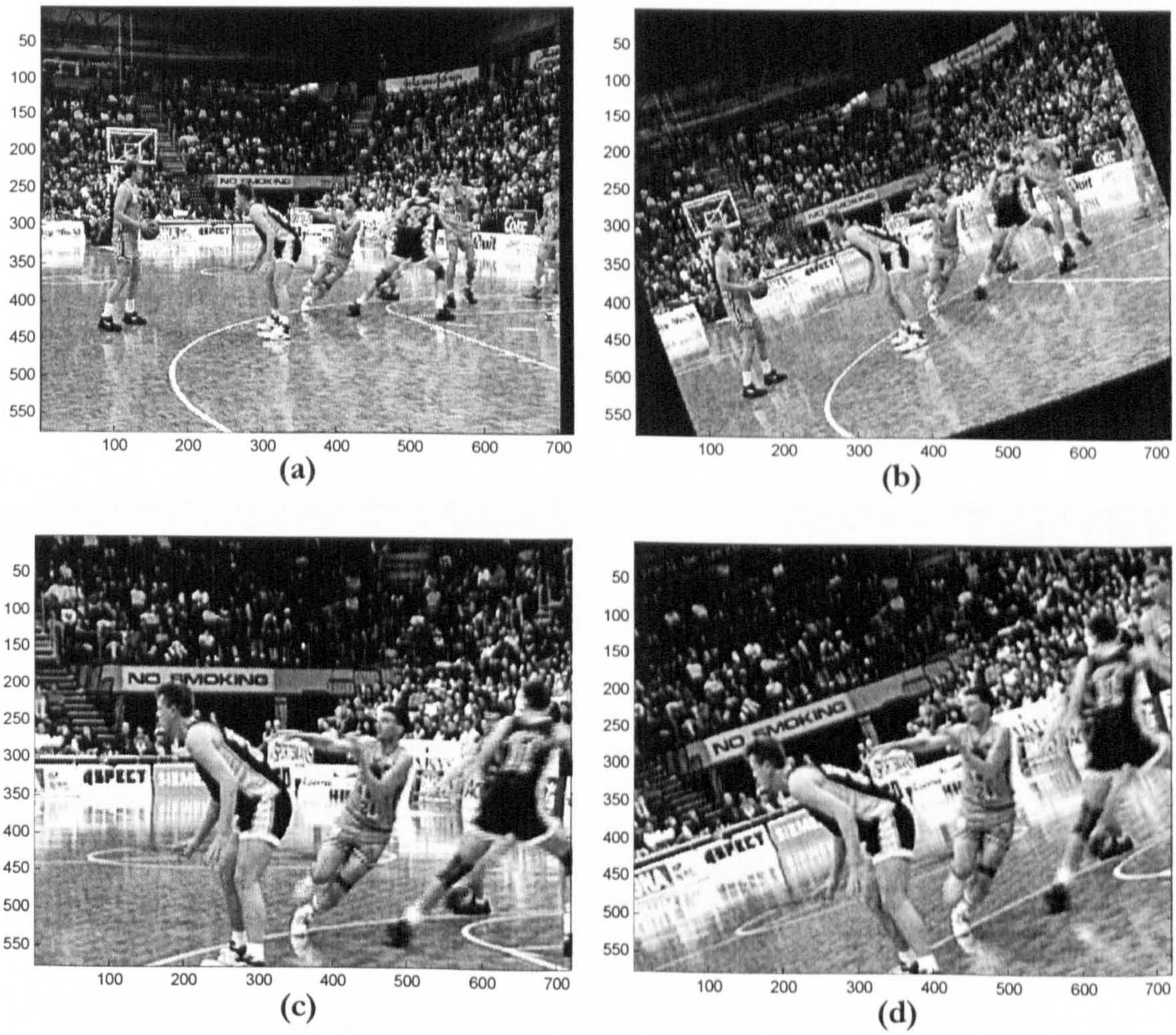


Figure 7-5 The effects of different geometrical attacks: **(a)** original basketball sequence, **(b)** 20^0 rotation, **(c)** 100% scaling and **(d)** 20^0 rotation combined with 100% scaling.

compression and shift combined with cropping and compression. The performance of the system for these attacks is presented below.

7.6.1 Scaling and Rotation Attack

These attacks are illustrated in **Figure 7-5** for "basketball" video sequence both as a discrete attack and as a combined attack. **Figure 7-7(a)** shows the performance of the system for different degrees of rotation, when n frames are averaged in order to improve the robustness of the system. Since the minimum watermarking segment is 25 frames, then $n \leq 25$. Compared with the $n = 1$ case, the cross-correlation peak for $n = 25$ is about four times larger. Similar results are presented in **Figure 7-7(b)** for scaling. **Figure 7-8(a)** and **Figure 7-8(b)** illustrates the system performance for $n = 25$ and different degrees of rotation

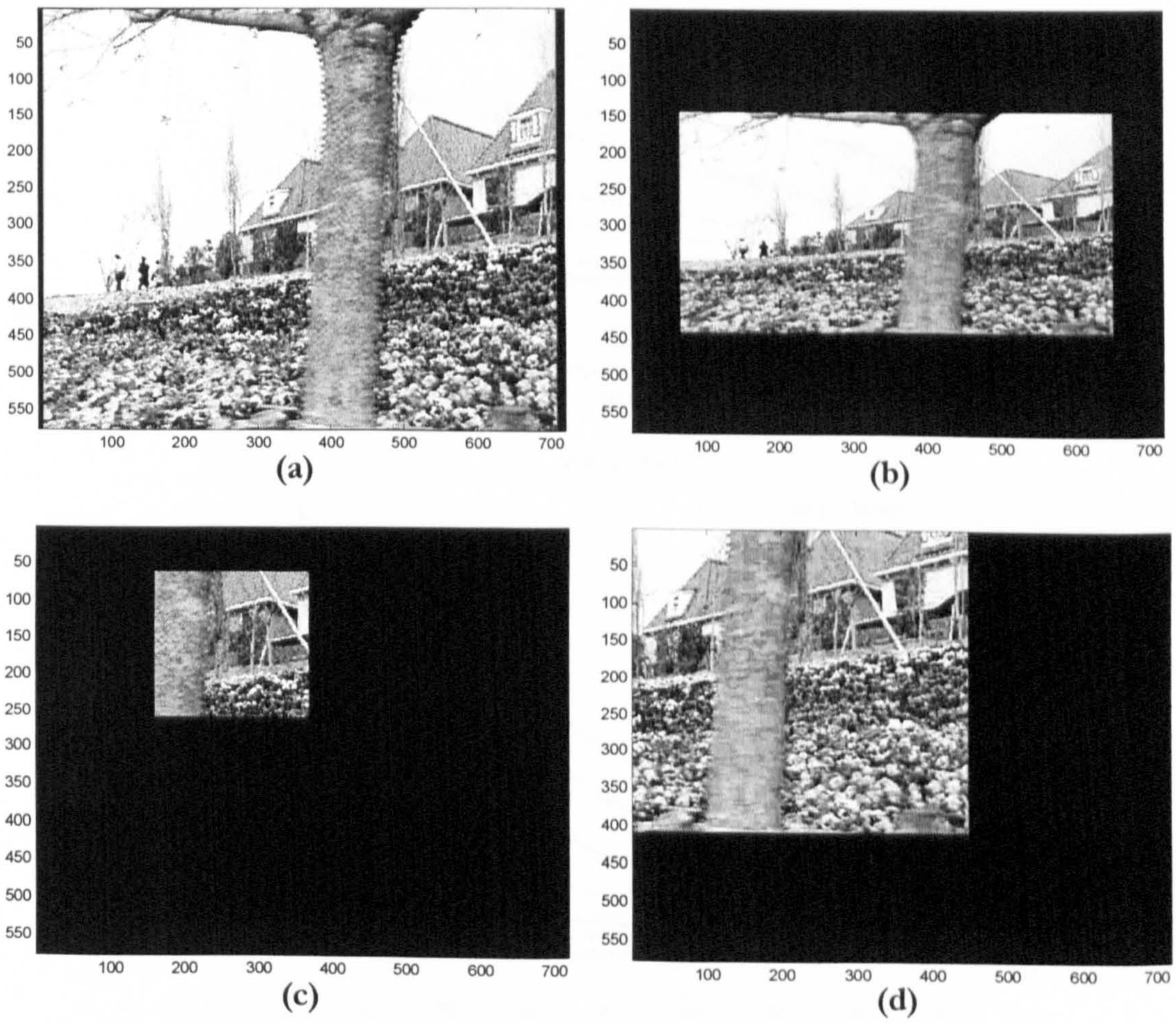
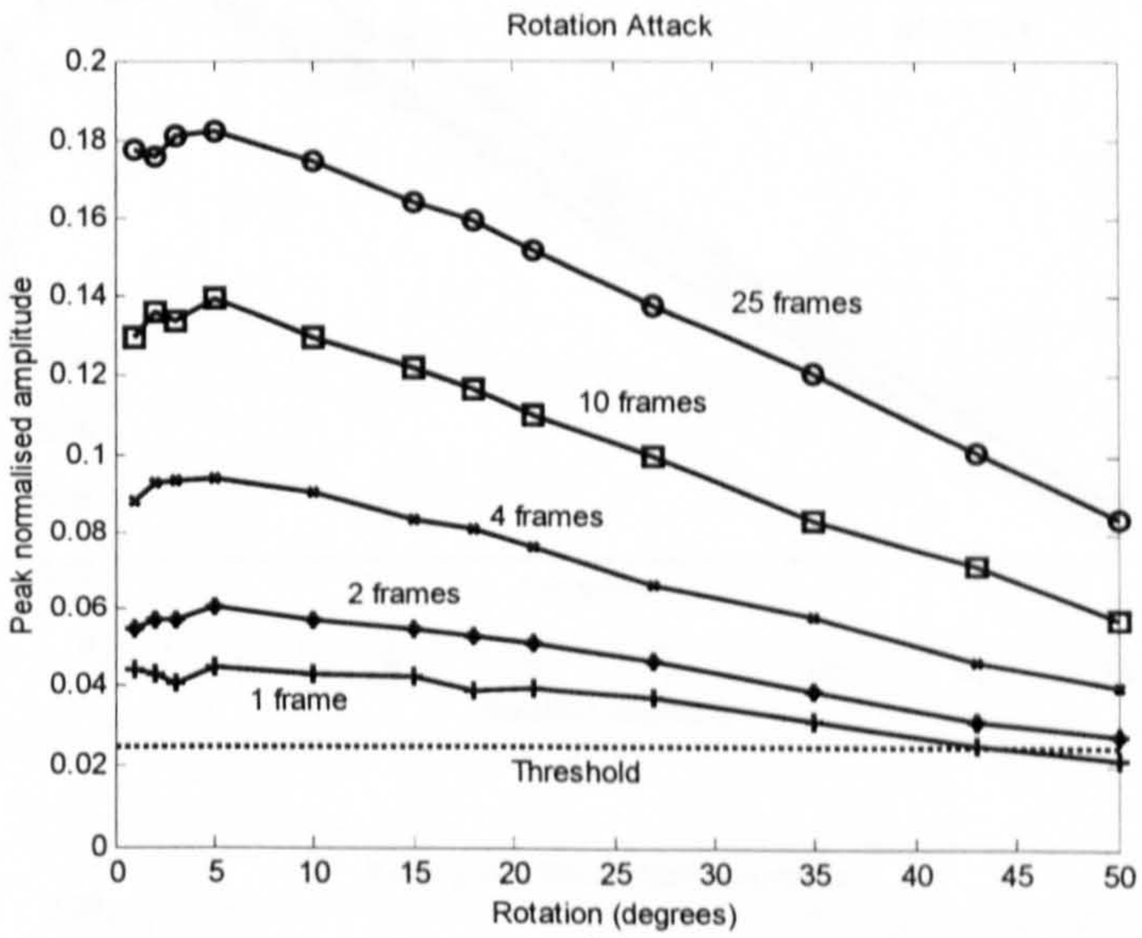


Figure 7-6 The effects of different attacks: **(a)** original flower sequence, **(b)** arbitrary scaling, from [576x720] to [300x600], **(c)** cropping [400,200,208,196] combined with shift [140,240], **(d)** shift [170,260] combined with 3Mbps MPEG2 compression.

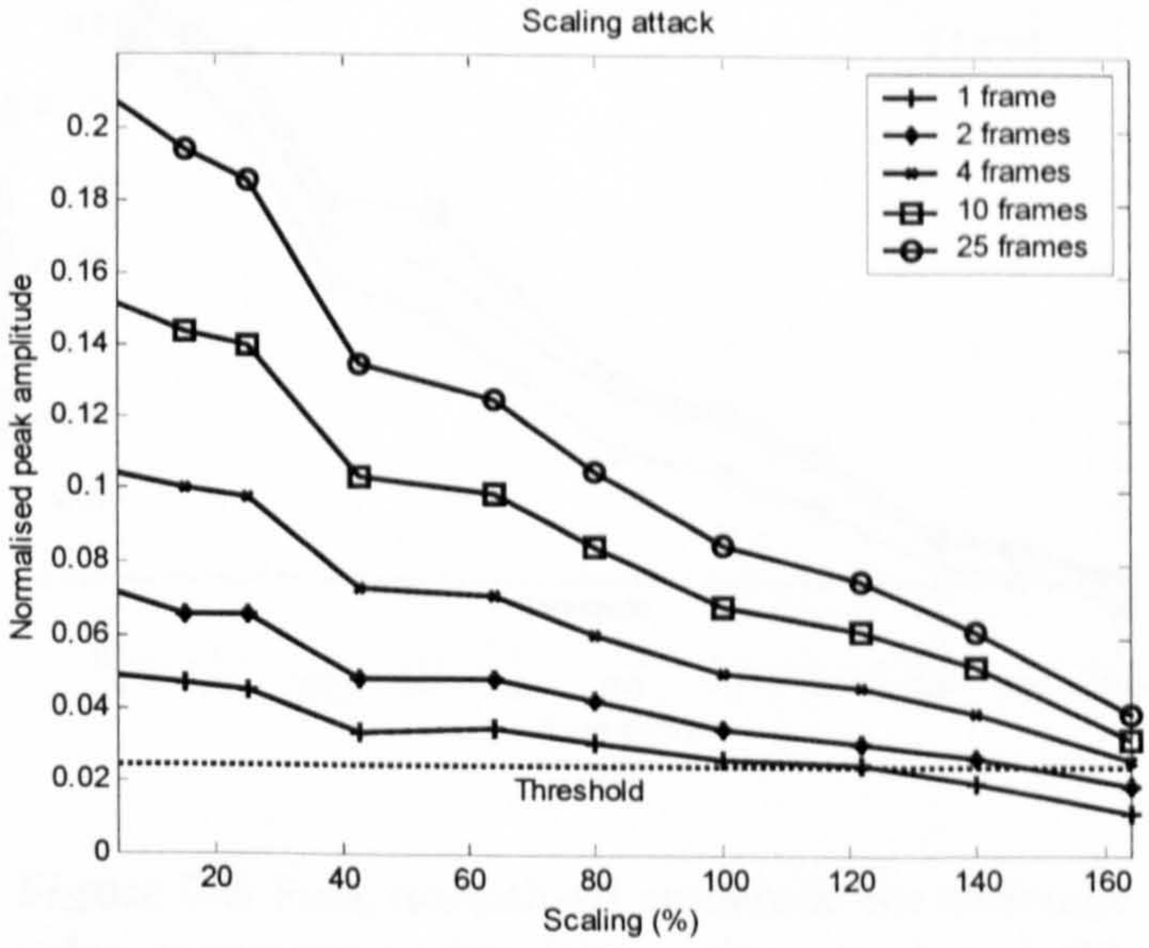
and respectively scaling for three typical video sequences. As expected, the worse case is the "flower" sequence. Finally, the combined attack (rotation plus scaling), is shown in **Figure 7-10** for the "basketball" sequence ($n = 25$).

As **Figure 7-8(a)** and **(b)** suggests, the system is invariant to any amount of rotation smaller than 70^0, and it can handle any degree of scaling up to 180%. The system is also capable to handle scaling up to –50% (i.e. smaller frames). Therefore the EBU recommendations [Cheveau et al, 2000] are exceeded for both rotation and scaling.

When rotation is combined with scaling, up to 120% scaling and up to 20^0 rotation can



(a)



(b)

Figure 7-7 Performance of the system when averaging frames for: **(a)** rotation and **(b)** scaling.

be tolerated, even for the "flower" sequence. All simulations assume a bilinear interpolation in the log-polar module. The experiments show that bilinear interpolation leads to a substantial performance increase (almost double) compared with a simple nearest neighbour interpolation. These two cases are illustrated in **Figure 7-9**. It is obvious from **Figure 7-9(a)** that the nearest neighbour interpolation leads to a much coarser result compared with the bilinear interpolation (**Figure 7-9(b)**).

Figure 7-8 Peak normalised amplitude for different video sequences under: **(a)** rotation attack and **(b)** scaling attack, when averaging 25 frames together.

7.6.2 Aspect Ratio Changing Attack

This attack is illustrated in **Figure 7-6(b)**. The difference between scaling and aspect ratio changing is that in the latter case two different scaling factors are present: one for horizontal scaling and a different one for vertical scaling. Using a log-log registration module, the system is invariant to arbitrary aspect ratio changes in the range –100% to 200% on both axes. The experiments show that this attack is much easier to handle than rotation combined with scaling. Again, bilinear interpolation performs much better compared with nearest neighbour interpolation (slightly more than double in this case).

7.6.3 Shifting and Cropping Attack

Shifting and cropping alone, or even the combined attack are not a posing a significant threat for the system, provided that at least 30% of the frame is still intact. These attacks are handled by the SPOMF registration module.

An illustration of a shifting and cropping attack is provided in **Figure 7-6(c)**. Even for this severe case, unlikely to happen in practice, the watermark can be recovered relatively easy. Again, in this case the EBU recommendations are largely exceeded.

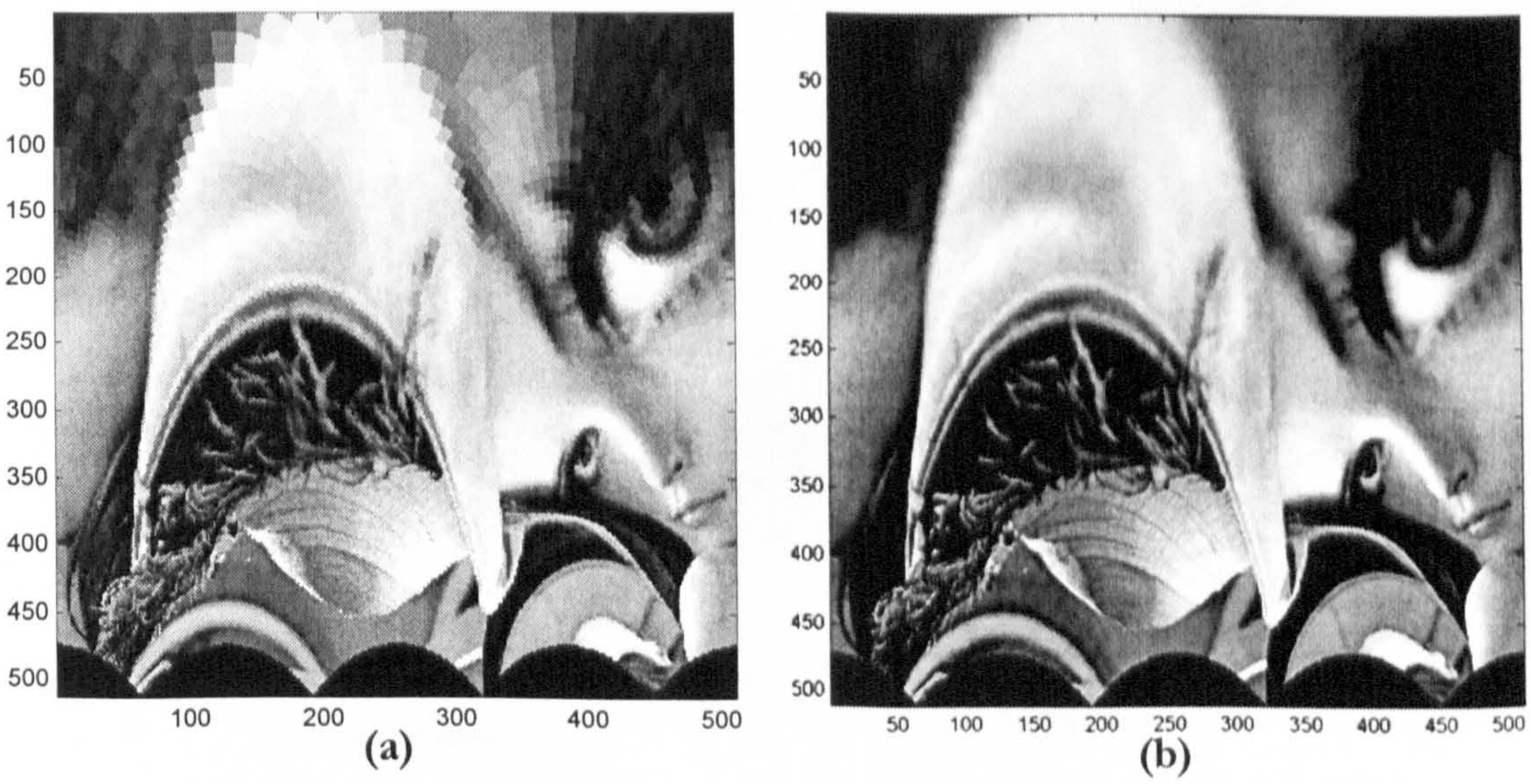


Figure 7-9 The log-polar map of image "Lena" for: **(a)** nearest neighbour interpolation and **(b)** bilinear interpolation.

7.6.4 Compression Attack

The system can cope very well with MPEG2 compression. The results under MPEG2 compression are presented in **Figure 7-11(a)**. One can see that the system can cope even with MPEG2 compression at 2Mbps, for all the test sequences involved.

Combined attacks like MPEG2 compression plus arbitrary frame shifts can be handled as long as the MPEG2 compression is at least 3-4Mbps (**Figure 7-11(b)**).

7.7 The False Detection Probability

A threshold value of 0.025 can be observed in each figure (**Figure 7-7**, **Figure 7-8**, **Figure 7-10**, **Figure 7-11**). This guarantees a false detection probability better than 10^{-8} when the correlation peak exceeds the threshold.

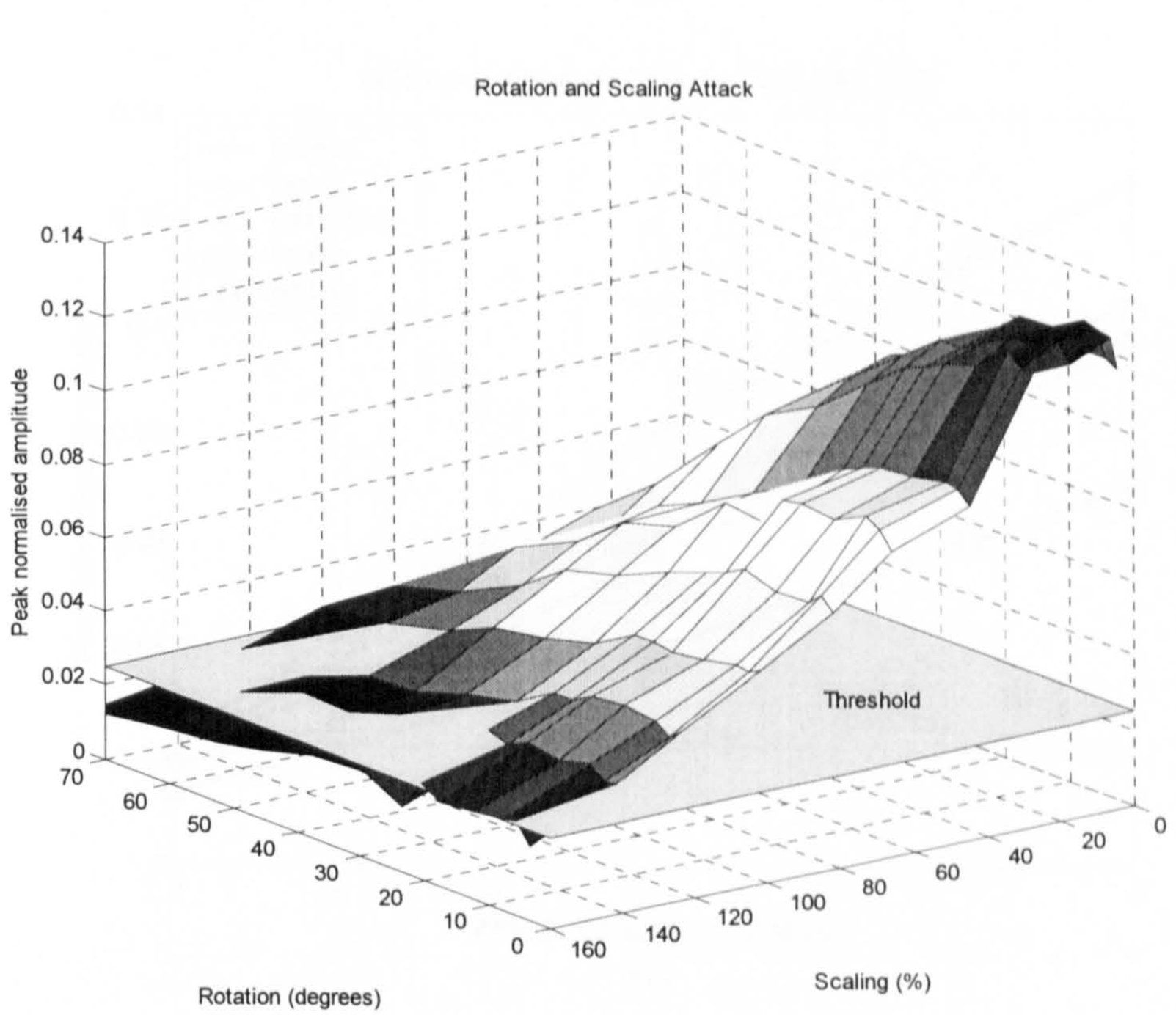


Figure 7-10 Performance of the system for rotation combined with scaling (25 frames, basketball video sequence).

The value was experimentally derived for a set of 3 test sequences and a wide range of scaling and rotation attacks: the pdf (probability distribution function) of the peaks was

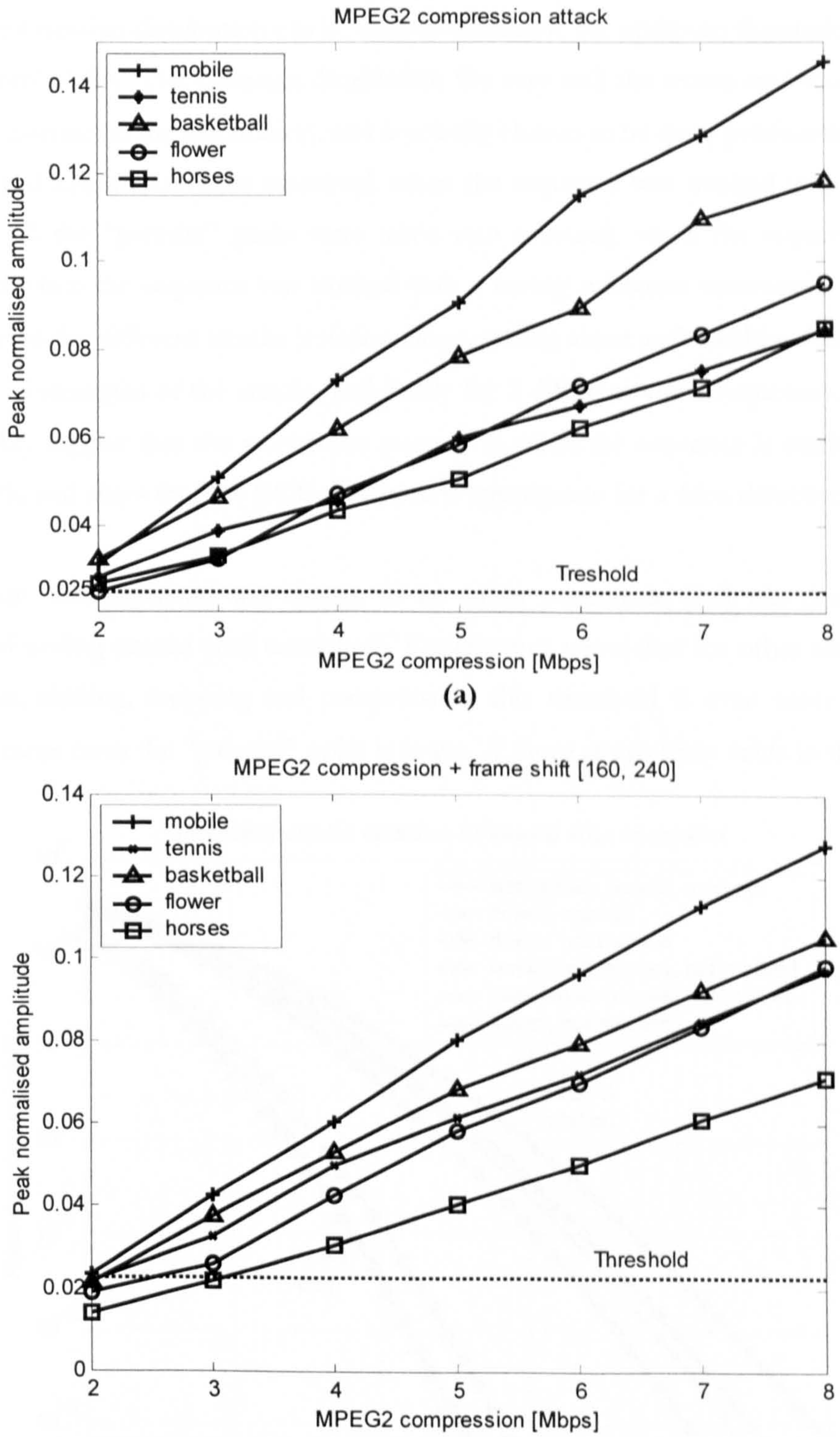


Figure 7-11 Compression attack: **(a)** MPEG2 compression, for different video sequences and **(b)** MPEG2 compression combined with spatial shift [160, 240].

computed for each case and the worse-case scenario determined. The resulting pdf is not Gaussian due to the large number of very small peak values, but by fitting a zero-mean Gaussian distribution with the same standard deviation as the experimentally determined pdf, the resulting Gaussian distribution can be used to determine the optimum threshold for a given false error probability. The Gaussian distribution fits very well the worse-case scenario pdf in the zone of interest (at the extremities), and is actually chosen to be quite pessimistic.

Several hypotheses were examined: when the sequence was marked with the correct watermark (all the "parasite" peaks were taken into account), when the sequence was not marked and when the sequence was marked with a wrong reference watermark. These cases were examined for different attacks (rotation alone, scaling alone and combined attacks) and a wide range of strengths of the attacks, and finally for 3 different video sequences. The results (**Figure 7-12**) suggest that the worse-case scenario is when the sequence is marked with the correct mark, and show that the 0.025 threshold is appropriate for a false detection probability of 10^{-8}.

Again, this threshold was chosen to be rather pessimistic, and was determined for rotation and scaling attacks (and combined). Experiments show that for other attacks (aspect ratio change, shifting, cropping and compression) this threshold is even more pessimistic, because in these cases the "parasite" noise is lower. A more appropriate value in these cases is

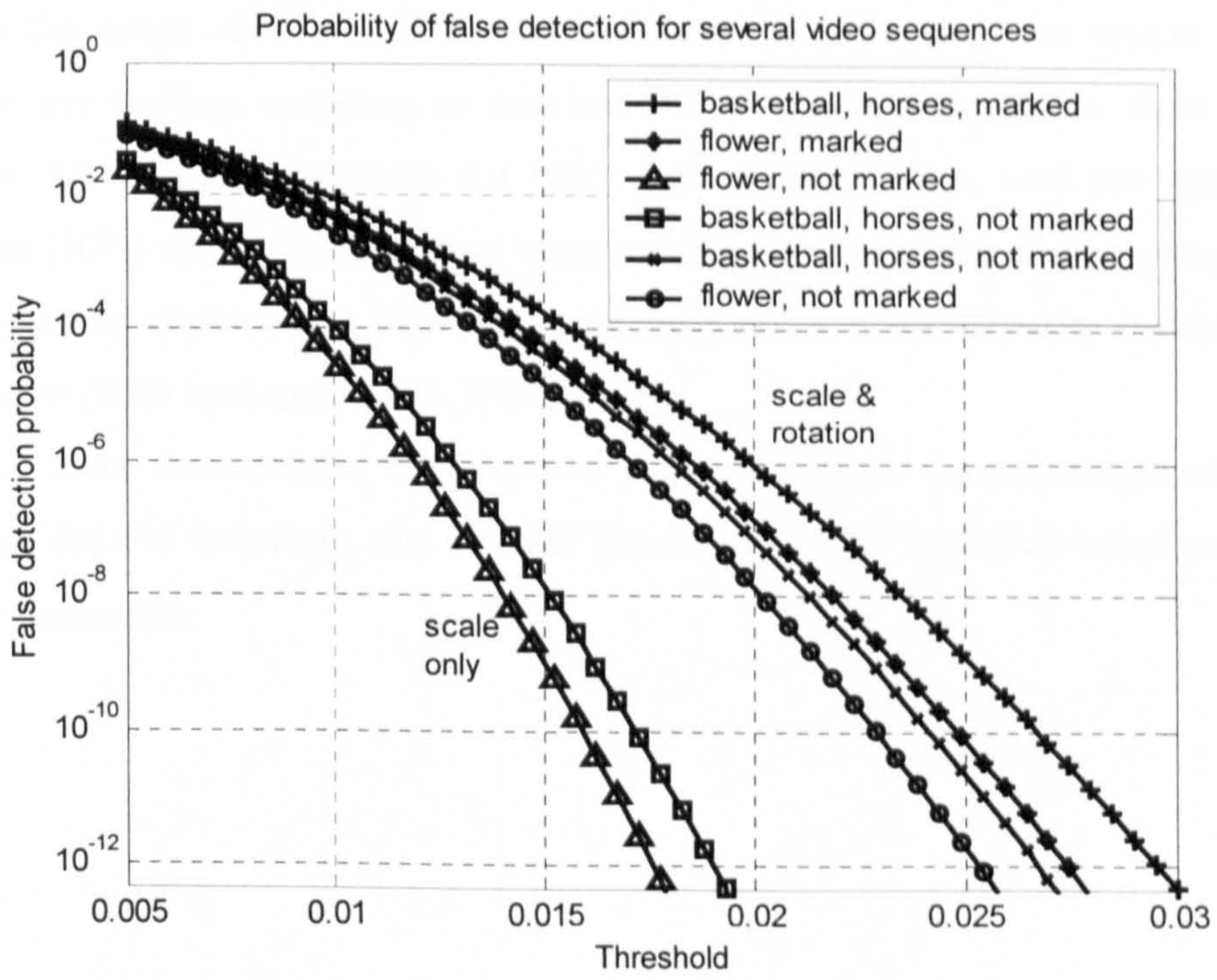


Figure 7-12 Threshold selection for a desired probability of false detection.

somewhere between 0.015-0.020.

7.8 Conclusions

Robustness to geometric attack is one of the most important requirements for a watermarking system. To satisfy this requirement an efficient approach based on the Fourier-Mellin transform and log-polar and log-log representations of the video frames has been developed. This is combined with the advantages of the DWT, HVS-based marking, and turbo coding to produce a very robust, high capacity video watermarking system.

With turbo coding, capacity can be as high as 1500 bits/frame (37.5 Kbps) even under severe cropping, and the system is invariant to a wide range of geometric attacks, such as scaling, rotation, aspect-ratio change, shifting, cropping and compression. It can also handle combined attacks, such as scaling/rotation, cropping/shifting, cropping/compression, shifting/compression, and cropping/shifting/compression. The search space is considerably reduced by using fast SPOMF-based cross-correlation.

For a false detection probability of 10^{-8}, the proposed system is invariant to scaling in the range –50% to 180%, invariant to rotation up to 70^0, and invariant to arbitrary aspect ratio changes in the range –100% to 200% on both axes. Furthermore, the system is virtually invariant to any shifting, cropping, or combined shifting and cropping. In these respects it exceeds the EBU recommendations for video watermarking. Also, very low quality JPEG compression (10%) can be handled even when combined with shifting and cropping, although this is not directly applicable to high quality uncompressed video. Capacity reduces to about 100 bits/frame (2500 bps) under 30% JPEG attack.

To achieve these results, the proposed system combines the advantages of geometric invariance of the FM transform, fast SPOMF processing, DWT and HVS-based marking, and turbo code protection.

Chapter 8

"Reading maketh a full man; conference a ready man; and writing an exact man."

Francis Bacon (1561-1626)

Conclusions & Further Work

Today virtually all multimedia production and distribution is digital. Therefore one can enjoy all the advantages of digital technology. The downside is that it becomes very easy to pirate copyrighted material as digital technology permits easy and unlimited copying without any loss in quality whatsoever. In order to combat this threat, one can use digital watermarking. The research carried out during this PhD and described in this thesis, addresses this difficult issue from the perspective of broadcast monitoring. This is just one of many possible applications of watermarking technology, as already shown in Chapter 1.

This chapter will conclude the work presented in this thesis and suggest further research directions. The complete list of author's publications can be also found here.

8.1 Conclusions

This thesis gradually described the building of a blind video watermarking system suitable for broadcast monitoring applications. Starting with a simple spatial domain technique, the system was gradually refined in order to meet the requirements imposed by broadcast monitoring. In the end, after passing several successive steps in order to improve the performance of the system, the result was a DWT-based, high capacity blind video watermarking system robust to a wide range of geometrical and non-geometrical attacks which meets and in some cases even exceeds (sometimes by far) the broadcast monitoring

requirements outlined by EBU [Cheveau et al, 2000]. This good performance was achieved by using a combination of communication techniques (spread spectrum, matched filtering, error correction coding), human perception (HVS models) and image processing techniques (image registration and pattern recognition).

Although EBU's recommendations specify a watermark capacity of 64 bps, the target of this research was to improve this figure as much as possible. This aim was achieved by regarding the watermarking channel as a communication channel which therefore can be protected by using FEC. Given the difficult nature of this channel, one needs powerful error correction in order to boost the performance of the system. This requirement is met by the Turbo codes, described in Chapter 4. Using this state-of-the-art FEC, the performance of the system increases significantly, as the comparative results provided in Chapter 5 and Chapter 6 show. Robustness to various attacks is a very important requirement in a watermarking system. In order to cope with this requirement, one would like to embed more energy into the host video. Unfortunately this is quite difficult since the invisibility of the mark has to be preserved. One solution for overcoming this aspect is to use HVS models which will accurately specify the maximum amount of modification for each pixel/coefficient while the invisibility constraint is still achieved. Therefore an investigation of the existing HVS models (mostly developed in the context of image compression) was carried out and the selected models were simplified, improved and adapted to the specific case of digital watermarking. Their efficiency and their benefits were presented in Chapter 3, Chapter 5 and Chapter 6.

One has the possibility of inserting a watermark either in spatial domain or in the transform domain. It is largely agreed that spatial domain is not as good as transform domain in terms of capacity/robustness, but has the advantage of simplicity. When the capacity is not a major issue, as in the case described in Chapter 7, this advantage is very important.

Speaking by transform domain watermarking, it is very important to choose a proper transform. Chapter 5 showed that the DCT is an important candidate in this respect. DCT domain watermarking offers much better performance compared with the spatial domain watermarking techniques described in Chapter 3. An important aspect of this research was to investigate the use of wavelet transform in digital watermarking. This investigation led to one of the main contributions of this thesis: developing a high capacity, robust, wavelet-based blind video watermarking system which takes advantage of the properties of the wavelet transform. The importance of using these advantages is well illustrated by the results presented in Chapter 6. The performance of the system is further improved by using HVS model and advanced FEC. Chapter 5 and Chapter 6 largely confirmed the superiority of the DWT against the DCT.

It has been clearly shown that DWT offers a much higher capacity/robustness report than DCT, even when using a much simpler HVS model.

Unfortunately the watermarking systems presented in Chapter 5 and Chapter 6 have one common negative characteristic: they cannot cope with geometrical attacks, unless an extensive search is performed in order re-synchronise the system. Using a 3-D sliding window correlator as suggested in Chapter 5 can solve this problem, but this approach has the major drawback of being very computationally expensive (difficult to use in a real time system), so it is preferably to employ other methods less demanding. Finding such methods and using them for blind video watermarking was another main challenge of this research, and represents the most important contribution of this thesis.

This problem was addressed in Chapter 7. This chapter shows that, by using image registration techniques (adapted to the specific of digital watermarking) in conjunction with a reference watermark, this problem can be solved much easier. The use of LPT/LLT in image registration was known from long time, but giving the blind nature of video watermarking, this technique could not be used in this case, since it requires the presence of the original image/video. The technique proposed in Chapter 7 shows that this problem can be solved by using a reference watermark in order to compensate for the unavailability of the original, and therefore achieving "blind registration". The reference watermark is embedded in spatial domain in order to keep the system as simple as possible. Using a separate reference watermark is advantageous because one can keep the high capacity of the system unchanged. Moreover the main watermark is embedded in wavelet domain, as suggested in Chapter 6, and therefore one can benefit from the advantages offered by the DWT transform. The net result is a DWT-based, high capacity blind video watermarking system robust to a wide range of geometrical and non-geometrical attacks. The performance of the final system is summarised in a tabular form, and compared with the EBU recommendations. This comparison, presented in **Table 8-1**, fully illustrates the performance and the capabilities of the proposed system.

Although the EBU recommendations do not specify (at this stage) any combined attacks, the system presented Chapter 7, can handle a varied range of combined attacks: shift combined with MPEG2 compression (>3Mbps MPEG2), shift combined with cropping (providing that approximately 30% from the original frame is still available), shift combined with cropping and compression, and even scaling combined with rotation (up to 120% scaling combined with up to 20^0 of rotation). Additionally, even if is not typical to video, the system can cope with very low quality JPEG compression.

Parameter	EBU Recommendations	Proposed System
GENERAL PARAMETERS OF THE SYSTEM		
Watermarking Minimum Segment (WMS)	1- 5sec	min 1sec
Data capacity	64bits/WMS	≥1200bits/WMS @ 2Mbps MPEG2
Probability for error-free payload per WMS	$>10^{-8}$	10^{-8}
False positive probability per WMS	$<10^{-8}$	$<10^{-8}$
Format of original and watermarked signals	ITU-R 601 (ITU-T BT.656)	ITU-R 601 (ITU-T BT.656)
Watermark recovery	Blind	Blind
ROBUSTNESS TO ATTACKS		
MPEG2 compression	2-6Mbps MPEG2	2-6Mbps MPEG2
VHS attack	YES	YES (easy)
Colour-space conversion	YES	Invariant
Collusion	YES	YES (easy)
Multiple Watermarks	YES	YES
Shift	up to 320x288	higher than 320x288
Scaling	desired: 200%, -50% best achieved: 140%, -70%	180%, -50%
Aspect-ratio conversion	16:9 ↔ 4:3	16:9 ↔ 4:3 (easy) 200%, -100%
Small rotation	up to 2^0	up to 2^0
Noticeable rotation	up to 10^0	up to 70^0
Small bend/shear	up to 2^0 (10^0)	NO
Cropping	minimum size 320x288	Even smaller than 200x200
Frame rate changing	24Hz ↔ 25Hz ↔ 30Hz	Invariant
Slow motion	3:1	Invariant
Combined attacks	NOT SPECIFIED	YES (wide range)

Table 8-1 The performance of the system compared with EBU's recommendations.

Starting from mid nineties quite a lot of companies are offering watermarking techniques for different market sectors. The international standardizing bodies and professional groups started several attempts to create some standards and recommendations for this potentially huge market. Although currently a watermarking standard is still missing, at least standardising bodies like EBU have issued some recommendations, making the first step in the direction of a standard. The direct result of the dissension within the watermarking world (from the industry's perspective), was that currently most of the digital technologies are not using watermarking techniques, opting instead in favour of cryptographic protection techniques like CSS (Content Scrambling System) for DVD and CI (Common Interface) for DVB. I spite of this, the watermarking research is a dynamic "market" with the research still going on in order to obtain better systems, proving once more that the interest in watermarking technology is still high. It is expected that sooner or later watermarking techniques will be probably incorporated in any (digital) multimedia system, especially since the CSS system was cracked pretty soon after the public release of the DVD and the DVB is confronted with a lot of piracy. Therefore the digital watermarking seems to have a very bright future ahead.

8.2 Further Work

In respect to EBU's recommendations, the system proposed in Chapter 7 has only one flaw: it cannot cope (directly) with geometric attacks like bending/shearing, unless an extensive search is performed. This is not a major issue, since the search space is quite limited (EBU recommends robustness to unnoticeable/small bending/shearing only). Obviously one would like to address this problem more efficiently in the future. On a related note, one could extend the system to cope with any affine transformation; in which case the bending/shearing would not be an issue anymore.

Choosing the best wavelet basis is very important. It would be very interesting to compare the performance of the system for different wavelet families. This could be difficult because of the lack of HVS models for most wavelet basis. Complex wavelets and their properties could offer a handy way of overcoming the affine transformations and maybe even non-affine transformations (StirMark like attacks). Using these wavelets could even improve

the capacity of the system, although at this moment there is no HVS model available for this wavelet.

As this thesis has shown, the HVS models play a central role in any watermarking system. The importance of a reasonable HVS model is paramount. The proposed system uses a very simple HVS model, and yet can achieve very good results. It is likely that using a more advanced HVS model will further improve the performance of the system and will reduce even more the watermark visibility. Using a JND-like model could result in optimum watermark embedding.

A more in depth analysis of the multiple watermarking case could be also performed, in order to asses exactly how many different watermarks can be inserted in the video sequence. This problem is obviously very closely related with the HVS model. It is likely that this will be the major limiting factor in having several watermarks on top of each other, since as the experiments show and literature agrees [Cheveau et al, 2000] different keys usually lead to near orthogonal watermarks and is not a major problem to have several watermarks on top of each other (from this point of view). This would evidently involve extensive (subjective) visibility tests in order to determine the impact of multiple watermarks on the visibility.

Although at the moment this is not a major concern (at least not from the broadcast monitoring perspective), the ownership deadlock problem could be sooner or later addressed. One way to solve this very difficult problem is to use One Way Hashing Functions and time stamps. The presence of a third party may be necessary in order to authenticate the keys. Generally the ownership deadlock is still an open problem and does not raise a significant interest in the watermarking community at this time.

Finally, although this is not specified at this time in the EBU recommendations, it would be very interesting to analyse the performance of the system under various other compression standards: MPEG4/DivX, MPEG7, MPEG21, JPEG2000.

8.3 List of Author's Publications

"Adding Robustness to Geometrical Attacks to a Wavelet Based, Blind Video Watermarking System", C. Serdean, M. Ambroze, M. Tomlinson, G. Wade, Proceedings of the IEEE International Conference on Multimedia and Expo – ICME 2002, Lausanne, Switzerland, 26-29 August 2002.

"DWT Based Video Watermarking for Copyright Protection, Invariant to Geometrical Attacks", C. Serdean, M. Ambroze, M. Tomlinson, G. Wade, Proceedings of the 3'rd International Symposium on Communication Systems, Networks & Digital Signal Processing – CSNDSP' 2002, Staffordshire, UK, 15-17 July 2002, pp. 312-315.

"Combating Geometrical Attacks in a DWT based Blind Video Watermarking System", C. Serdean, M. Ambroze, M. Tomlinson and G. Wade. Proceedings of the 4'th EURASIP-IEEE International Symposium on Video/Image Processing & Multimedia Communications – VIPromCom 2002, Zadar, Croatia, 16-19 June 2002, pp. 263-266.

"Protecting Intellectual Rights: Digital WM in the Wavelet Domain", C. Serdean, M. Tomlinson, G. Wade & M. Ambroze, Proceedings of the IEEE International Workshop "Trends & Recent Achievements in Information Technology", Cluj-Napoca, Romania, 16-18 May 2002, pp.70-77.

"DWT Based High Capacity Blind Video Watermarking, Invariant to Geometrical Attacks", C. Serdean, A. Ambroze, M. Tomlinson and G. Wade, Accepted for publication in IEE Proceedings Vision, Image and Signal Processing, Submitted December 2001 (in press).

"Turbo Code Protection of a Video Watermarking Channel", A. Ambroze, G. Wade, C. Serdean, M. Tomlinson, J. Stander and M. Borda. Published in IEE Vision, Image and Signal Processing, Vol.148, No.1, February 2001, pp. 54-58.

"Watermarking Uncompressed Video: an Overview", G. Wade, C. Serdean, A. Ambroze, M. Borda & I. Nafornita. Proceedings of the IEEE Symposium on Electronics and Telecommunications, 'Etc. 2000', 23-24 November 2000, Timisoara, Romania, Vol.1, pp. 2-15. (Invited Paper)

Acknowledgement

Parts of the work described during Chapter 5 were carried out in collaboration with Dr. Adrian Ambroze. The main contribution of the author was related with watermark embedding (HVS, modulation techniques, system improvements) and watermark recovery (2-D sliding correlator, visibility tests). Dr. Ambroze was in charge with channel protection (pdf of the channel, Turbo coding, performance assessment) and attack characterisation. Also most of the work related with watermark recovery (3-D sliding correlator, several performance tests) was carried out in parallel with Dr. Ambroze.

The author also wants to thank Dr. Ambroze for his assistance during the many visibility/robustness tests carried out and for supplying him with several performance diagrams (**Figure 5-13 – Figure 5-19**) which saved allot of computation time. The author gratefully acknowledges the above contributions and wishes to thank Dr. Ambroze for a fruitful collaboration.

References

A

[Ahumada et al, 1992] A.J. Ahumada Jr. and H.A. Peterson, "Luminance-model-based DCT Quantization for Colour Image Compression", Human Vision, Visual Processing and Digital Display III, Proceedings of the SPIE, B.E. Rogowitz Editor, Vol. 1666, pp. 365-374, 1992.

[Ambroze, 2000] M.A. Ambroze, "On Turbo Codes and Other Concatenated Schemes in Communication Systems", PhD Thesis, University of Plymouth, September 2000.

[Ambroze et al, 2001] M.A. Ambroze, G.J. Wade, C.V. Serdean, M. Tomlinson, J. Stander and M. Borda, "Turbo Code Protection of Video Watermark Channel", IEE Proc. Vision, Image and Signal Processing, Vol.148, No.1, pp. 54-58, February 2001.

[Antonini et al, 1992] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, "Image Coding Using the Wavelet Transform", In IEEE Trans. Image Processing, Vol. 2, pp. 205-220, April 1992.

B

[Balado et al, 2001] Félix Balado and Fernando Pérez-González, "Coding at the Sample Level for Data Hiding: Turbo and Concatenated Codes", In Ping Wah Wong and Edward J. Delp editors, Security and Watermarking of Multimedia Contents III, Proc. of SPIE, Vol. 4314, pp. 532-543, San Jose, USA, January 2001.

[Barbulescu, 1996] A.S. Barbulescu, "Iterative Decoding of Turbo Codes and Other Concatenated Codes", PhD Thesis, University of South Australia, February 1996.

[Barni et al, 1998-1] M. Barni, F. Bartolini, A. Piva and F. Rigacci, "Statistical Modelling of Full Frame DCT Coefficients", In Proceedings of EUSIPCO'98, 9'th European Signal Processing Conference, pp. 1513-1516, Rhodes, Greece, 8-11 September, 1998.

[Barni et al, 1998-2] M. Barni, F. Bartolini, V. Cappellini, A. Piva and F. Rigacci, "A MAP Identification Criterion for DCT-based Watermarking", Proc. European Signal Processing Conf. EUSIPCO 98, Island of Rhodes, Greece, 8-11 September, 1998.

[Barni et al, 1998-3] Mauro Barni, Franco Bartolini, Vito Cappellini and Alessandro Piva, "A DCT-domain System for Robust Image Watermarking", Signal Processing, Vol. 66, No. 3, pp. 357-372, May 1998.

[Barni et al, 1999-1] M. Barni, F. Bartolini, A. De Rosa and A. Piva, "Capacity of the Watermark Channel: How Many Bits Can Be Hidden Within a Digital Image?" Proc. SPIE, Vol. 3657, pp. 437-448, San Jose, CA, January 1999.

[Barni et al, 1999-2] M. Barni, F. Bartolini, V. Cappellini, A. Lippi and A. Piva, "A DWT-based Technique for Spatio-frequency Masking of Digital Signatures", Proc. SPIE/IS&T Int. Conference on Security and Watermarking of Multimedia Contents, Vol. 3657, San Jose, CA, 25-27 January, 1999.

[Bartolini et al, 1998] F. Bartolini, M. Barni, V. Cappelini and A. Piva, "Mask Building for Perceptually Hiding Frequency Embedded Watermarks", Proc. of 5'th IEEE Int. Conference on Image Processing ICIP98, Vol. 1, pp. 450-454, Chicago, Illinois, USA, 4-7 October, 1998.

[Baudry et al, 2001] S. Baudry, B. Sankur, J.F. Delaigle, B. Macq and H. Maitre, "Analyses of Error Correction Strategies for Typical Communication Channels in Watermarking", Signal Processing, Vol. 81, No. 6, pp. 1239-1250, June 2001.

[op de Beeck et al, 2001] M.J. op de Beeck, J. Haitsma and A.A.C. Kalker, Koninklijke Philips Electronics N.V., "Watermark Detection", WIPO, International Patent WO 01/24113 A1, 5 April, 2001.

[Bender et al, 1995] W. Bender, D. Gruhl and N. Morimoto, "Techniques for Data Hiding", In Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases III, Vol. 2420, pp. 164-173, San Jose, CA, February 1995.

[Berrou et al, 1993] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo-Codes", In Proc. of the IEEE Int. Conf. on Communications, pp. 1064-1070, Geneva, Suisse, May 1993.

[Boland et al, 1995] F.M. Boland, J.J.K. O'Ruanaidh and C. Dautzenberg, "Watermarking Digital Images for Copyright Protection", In Proceedings of the International Conference on Image Processing and its Applications, pp. 321-326, Edinburgh, Scotland, July 1995.

[Braudaway, 1999] Gordon Braudaway, "Recovering Invisible Image Watermarks from Images Distorted by the "StirMark" Algorithm", IBM Research Report, RC21396, 1999.

[Burke-Hubbard, 1998] Barbara Burke Hubbard, "The World According to Wavelets: The Story of a Mathematical Technique in the Making", Second Edition, A.K. Peters, Wellesley, Massachusetts, 1998.

[Busch et al, 1999] C. Busch, W. Funk and S. Wolthusen, "Digital Watermarking: From Concepts to Real-Time Video Applications", IEEE Computer Graphics and Applications, pp. 25-35, Jan/Feb 1999.

C

[Carlson et al, 1980] C.R. Carlson and R.W. Cohen, "A Simple Psychophysical Model for Predicting the Visibility of Displayed Information", In Proceedings of the Society for Information Display, Vol. 21, No.3, pp. 229-246, 1980.

[Casasent et al, 1976-1] D. Casasent and D. Psaltis, "Scale-invariant Optical Correlation Using Mellin Transforms", Optics Communications, Vol. 17, pp. 59-63, April 1976.

[Casasent et al, 1976-2] David Casasent and Demetri Psaltis, "Position, Rotation and Scale Invariant Optical Correlation", Applied Optics, Vol. 15, No. 7, pp. 1795-1799, July 1976.

[Chen et al, 1994] Q.S. Chen, J. Defrise and F. Deconinck, "Symmetric Phase-only Matched Filtering of Fourier-Mellin Transforms for Image Registration and Recognition", Trans. IEEE, Pattern Analysis and Machine Intelligence, Vol. 16, No. 12, pp. 1156-1168, December 1994.

[Cheng et al, 1998] L. Cheng and J. Robinson, "Dealing with Speed and Robustness Issues for Video-based Registration on a Wearable Computing Platform", In Proceedings of the 2'nd International Symposium on Wearable Computers, pp. 84-91, IEEE CS Press, Los Alamitos, California, USA, October 1998.

[Cheveau et al, 2000] L. Cheveau, E. Goray and R. Salmon, "Watermarking – Summary Results of EBU Tests", EBU Technical Review, No. 282, March 2001.

[Chou et al, 1995] C. Chou and Y. Li, "A Perceptually Tuned Sub-band Image Coder Based on the Measure of Just-Noticeable-Distortion Profile", IEEE Trans. on Circuits and Systems for Video Technology, Vol. 5, No. 6, pp. 467-476, December 1995.

[Chou et al, 1996] C.H. Chou and C.W. Chen, "A Perceptually Optimized 3-D Sub-band Codec for Video Communication Over Wireless Channels", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 6, No. 2, pp. 143-156, April 1996.

[Couch, 1987] Leon W. Couch II, "Digital and Analog Communication Systems", Second Edition, Macmillan Publishing Company, 1997.

[Cox et al, 1995] I.J. Cox, J. Kilian, T. Leighton and T. Shamoon, "Secure Spread Spectrum Communication for Multimedia", NEC Research Institute, Technical Report 95-10, 1995.

[Cox et al, 1996] I.J. Cox, J. Kilian, T. Leighton and T. Shamoon, "A Secure, Robust Watermark for Multimedia", In Proceedings of the 1'st Int. Workshop In Information Hiding, Cambridge, UK, May 1996, R. Anderson editor, Lecture Notes in Computer Science, Vol. 1174, pp. 185-206, Springer-Verlag, 1996.

[Cox et al, 1997] I.J. Cox, M.L. Miller, "A Review of Watermarking and the Importance of Perceptual Modelling", Proceedings of SPIE, Electronic Imaging '97, Vol. 3016, pp. 92-99, San Jose, CA, USA, February 1997.

[Cox et al, 1998] I.J. Cox and J.P. Linnartz, "Some General Methods for Tampering With Watermarks", IEEE Journal on Selected Areas in Communications, Special Issue on Copyright and Privacy Protection, Vol. 16, No. 4, pp. 587-593, May 1998.

[Cox et al, 1999] I.J. Cox, M.L. Miller and A.L. McKellips, "Watermarking as Communications with Side Information", Proc. IEEE, Vol. 87, No. 7, pp. 1127-1141, 1999.

[Craver et al, 1996] Scott Craver, Nasir Memon, Boon-Lock Yeo and Minerva Yeung, "Can Invisible Watermarks Resolve Rightful Ownerships?", IBM Research Report, Technical Report, July 1996.

[Craver et al, 1997] S. Craver, N. Memon, B.-L. Yeo and M. Yeung, "On the Invertibility of Invisible Watermarking Techniques", In Proceedings of the IEEE International Conference on Image Processing ICIP'97, Vol. 1, pp. 540-543, Santa Barbara, CA, USA, October 1997.

[Craver et al, 1998] S. Craver, N. Memon, B. Yeo and M. Yeung, "Resolving Rightful Ownerships With Invisible Watermarking Techniques: Limitations, Attacks and Implications", IEEE Journal on Selected Areas in Communications, Vol. 16, No. 4, pp. 573-586, May 1998.

D

[Deguillaume et al, 1999] F. Deguillaume, G. Csurka, J.J.K. O'Ruanaidh and T. Pun, "Robust 3D-DFT Video Watermarking", In IS&T/SPIE Electronic Imaging '99, Session: Security and Watermarking of Multimedia Contents, San Jose, CA, USA, January 1999.

[Delaigle, 1998] J.F. Delaigle, C. De Vleeschouwer and B. Macq, "Watermarking Algorithm Based on a Human Visual Model", Signal Processing, Vol. 66, No. 3, pp. 319-335, May 1998.

[Delaigle, 2000] Jean-François Delaigle, "Protection of Intellectual Property by Perceptual Watermarking", PhD Thesis, Université Catholique de Louvain, Louvain-la-Neuve, Belgique, November 2000.

[Depovere et al, 1998] G. Depovere, T. Kalker and J.P. Linnartz, "Improved Watermark Detection Reliability Using Filtering Before Correlation", In Proc. IEEE Int. Conference on Image Processing ICIP'98, Vol. I, pp. 430-434, Chicago, Illinois, USA, October 1998.

[Dittmann et al, 2000] J. Dittmann, T. Fiebig, and R. Steinmetz, "New Approach for Transformation-invariant Image and Video Watermarking in the Spatial Domain: Self-spanning Patterns (SSP)", In Ping Wah Wong and Edward J. Delp, editors, Electronic Imaging 2001.

[Dolinar et al, 1998] S. Dolinar, D. Divsalar and F. Pollara, "Code performance as a function of blocksize", The Telecommunications and Mission Operations Progress Report 42-133, January-March 1998, Jet Propulsion Laboratory, Pasadena, California, pp. 1-23, 15 May, 1998.

[Dugad et al, 1998] R. Dugad, K. Ratakenda and N. Ahuja, "A New Wavelet-based Scheme for Watermarking Images", Proc. IEEE, Int. Conf. on Image Processing ICIP'98, Vol. 2, TA10.07, Chicago, USA, October 1998.

E

[Eckert, 1998] Michael P. Eckert and Andrew P. Bradley, "Perceptual Models Applied to Still Image Compression", Signal Processing, Vol. 70, No. 3, pp. 177-200, November 1998.

F

[Fei et al, 2001] C. Fei, D. Kundur and R.H. Kwong, "The Choice of Watermark Domain in the Presence of Compression", In Proc. IEEE Int. Conf. on Information Technology: Coding and Computing, pp. 79-84, April 2001.

G

[Girod, 1989] B. Girod, "The Information Theoretical Significance of Spatial and Temporal Masking in Video Signals", In Proc. of the SPIE Human Vision, Visual Processing, and Digital Display, Vol. 1077, pp. 178-187, Los Angeles, CA, 1989.

H

[Hartung et al, 1996] F. Hartung and B. Girod, "Digital Watermarking of Raw and Compressed Video", Proc. European EOS/SPIE Symposium on Advanced Imaging and Network Technologies, Berlin, Germany, October 1996.

[Hartung et al, 1998] F. Hartung and B. Girod, "Watermarking of Uncompressed and Compressed Video", Signal Processing, Vol. 66, No. 3, pp. 283-301, 1998.

[Hartung et al, 1999-1] F. Hartung, J.K. Su and B. Girod, "Spread Spectrum Watermarking: Malicious Attacks and Counterattacks", Electronic Imaging '99, In Security and Watermarking of Multimedia Contents, Vol. 3657, pp.147-158, San Jose, USA, 24-29 January, 1999.

[Hartung et al, 1999-2] F. Hartung and M. Kutter, "Multimedia Watermarking Techniques", Proceedings of the IEEE: Special Issue on Protection of Multimedia Content, Vol. 87, pp. 1079-1107, July 1999.

[Hernandez et al, 1998-1] Juan R. Hernández, Fernando Pérez-González and José M. Rodríguez, "The Impact of Channel Coding on the Performance of Spatial Watermarking for Copyright Protection", In Proc. ICASSP'98, Vol. 5, pp. 2973-2976, Seattle, USA, May 1998.

[Hernandez et al, 1998-2] Juan R. Hernández, Fernando Pérez-González and Martín Amado, "Improving DCT-domain Watermark Extraction Using Generalized Gaussian Models" In Proc. of the COST 254 Int. Workshop on Intelligent Communications and Multimedia Terminals, pp. 23-26, Ljubljana, Slovenia, November 1998.

[Hernandez et al, 1999-1] Juan R. Hernández, Martín Amado and Fernando Pérez-González, "Novel Detector Structures for Watermark Extraction in the DCT Domain", In Proc. Workshop on Nonlinear Signal and Image Processing, Antalya, Turkey, June 1999.

[Hernandez et al, 1999-2] Juan R. Hernández and Fernando Pérez-González, "Statistical Analysis of Watermarking Schemes for Copyright Protection of Images", Proceedings of the IEEE, Special Issue on Identification and Protection of Multimedia Information, Vol. 87, No. 7, pp. 1142-1166, July 1999.

[Hernandez et al, 2000-1] Juan R. Hernández, Martín Amado and Fernando Pérez-González, "DCT-domain Watermarking Techniques for Still Images: Detector Performance Analysis and a New Structure", IEEE Trans. on Image Processing, Special Issue on Image and Video Processing for Digital Libraries, Vol. 9, No. 1, pp. 55-68, January 2000.

[Hernandez et al, 2000-2] J.R. Hernandez, J.F. Delaigle and B. Macq, "Improving Data Hiding by Using Convolutional Codes and Soft-decision Decoding", In Ping Pah Wong, E.J. Delp Editors, Security and Watermarking of Multimedia Contents, SPIE, Vol. 3971, San-Jose, CA, USA, January 2000.

[Hill et al, 1999] Lyndon Hill and T. Vlachos, "On the Estimation of Global Motion using Phase Correlation for Broadcast Applications", IEE International Conference on Image Processing and it's Applications IPA'99, pp. 721-725, Manchester, 1999.

[Horner et al, 1984] J.L. Horner and P.D. Gianino, "Phase-only Matched Filtering", Applied Optics, Vol. 23, No. 6, pp. 812-816, 1984.

[Huang et al, 2000] C.H. Huang and Ja-Ling Wu, "A Watermark Optimization Technique Based on Genetic Algorithms", In Proc. SPIE, Security and Watermarking of Multimedia Contents II, Vol. 3971, 24-26 January, 2000.

I

[Inoue et al, 1998] Hisashi Inoue, Akio Miyazaki and Takashi Katsura, "An Image Watermarking Method Based on the Wavelet Transform", Proceedings of 6'th International Conference on Image Processing ICIP'99, pp. 296-300, Vol. 1, Kobe, Japan, October 1999.

J

[Jain, 1989] Anil K. Jain, "Fundamentals of Digital Image Processing", Prentice Hall Inc., Englewood Cliffs, NJ, 1989.

[Jayant et al, 1993-1] N. Jayant, J. Johnston and R. Safranek, "Signal Compression Based on Models of Human Perception", Proc. IEEE, Vol. 81, No. 10, pp. 1385-1422, October 1993.

[Jayant et al, 1993-2] N. Jayant, J. Johnston and R. Safranek, "Perceptual Coding of Images", Proc. Society of Photo Instrumentation Engineering, SPIE, Vol. 1913, pp. 168-178, 1993.

[Jayawardena et al, 2000] A. Jayawardena, B. Murison and P. Lenders, "Embedding Multiresolution Binary Images Into Multiresolution Watermark Channels in Wavelet Domain", Visual Communications and Image Processing VCIP'2000, Proceedings of SPIE, Vol. 4067, Perth, Australia, 21-23 June, 2000.

[Johnson et al, 1998] N.F. Johnson and S. Jajodia, "Exploring Steganography: Seeing the Unseen", IEEE Computer, Vol. 31, No. 2, pp.26-34, February 1998.

K

[Kalker et al, 1999-1] A.A.C. Kalker, G. Depovere, J. Haitsma and M. Maes, "A Video Watermarking System for Broadcast Monitoring", Proceedings of the SPIE Security and Watermarking of Multimedia Contents, Vol. 3657, pp. 103-112, San Jose, 25-27 January, 1999.

[Kalker et al, 1999-2] A.A.C. Kalker, J. Haitsma, G. Depovere and J.P.M. Linnartz, Koninklijke Philips Electronics N.V., "Watermark Detection", WIPO, International Patent WO 99/45706 A2, 10 September, 1999.

[Kalker, 2000] A.A.C. Kalker, "Omnipresent and Invisible Digital Video Watermarking", Philips Research Password, No.5, October 2000.

[Kalker, 2001] A.A.C. Kalker, "Considerations on Watermarking Security", Proc. IEEE Workshop on Multimedia Signal Processing MMSP'2001, Cannes, France, 3-5 October, 2001.

[Kilian et al, 1999] Joe Kilian, F. Thomson Leighton, Lesley R. Matheson, Talal G. Shamoon, Robert E. Tarjan and Francis Zane, "Resistance of Digital Fingerprints to Collusional Attacks", Proceedings of IEEE International Symposium on Information Theory ISIT'98, pp. 271, Cambridge, MA, 16-21 August, 1998.

[Kim et al, 1999] S. Kim, S. Suthaharan, H.K. Lee and K.R. Rao, "Image Watermarking Scheme Using Visual Model and BN Distribution", Electronic Letters, Vol. 35, No.3, pp. 212-213, February 1999.

[Kingsbury, 1997] N. Kingsbury and N. Magarey, "Wavelet Transforms in Image Processing", In Signal Processing and Prediction I, EURASIP, ICT Press, pp. 23-34, Prague, Czech Republic, 1997.

[Kingsbury, 1998] N. Kingsbury, "The Dual-tree Complex Wavelet Transform: A New Technique for Shift Invariance and Directional Filters", In Proc. 8'th IEEE DSP Workshop, Paper No. 86, Bryce Canyon, UT, USA, 9-12 August, 1998.

[Kingsbury, 1999] N. Kingsbury, "Image processing with complex wavelets", In Phil. Trans. Royal Society London, Set A, Special issue for the discussion meeting on "Wavelets: the key to intermittent information?", Vol. 357, No. 1760, pp. 2543-2560, September 1999.

[Knuth, 1981] D.E. Knuth, "The Art of Computer Programming: Semi-numerical Algorithms", Vol. 2, Second Edition, Addison-Wesley, Reading, Massachusetts, 1981.

[Kobayashi, 1997] Mei Kobayashi, "Digital Watermarking: Historical Roots", Technical report, IBM Research, Tokyo Research Laboratory, April 1997.

[Kuglin et al, 1975] C.D. Kuglin and D.C. Hines, "The Phase Correlation Image Alignment Method", In Proceedings of the IEEE, International Conference on Cybernetics and Society, pp. 163-175, New York, 1975.

[Kundur et al, 1997] Deepa Kundur and Dimitrios Hatzinakos, "A Robust Digital Image Watermarking Method Using Wavelet-based Fusion", In International Conference on Image Processing ICIP'97, pp. 544-547, Santa Barbara, California, USA, October 1997.

[Kundur et al, 1998] Deepa Kundur and Dimitrios Hatzinakos, "Digital Watermarking Using Multiresolution Wavelet Decomposition", In Proc. of the International Conference on Acoustic, Speech and Signal Processing ICASP'98, Vol. 5, pp. 2969-2972, Seattle, Washington, USA, May 1998.

[Kundur, 1999] D. Kundur, "Multiresolution Digital Watermarking: Algorithms and Implications for Multimedia Signals", PhD Thesis, Dept. of Electrical & Computer Engineering, University of Toronto, August 1999.

[Kutter, 1998] M. Kutter, "Watermarking Resisting to Translation, Rotation and Scaling", Proceedings of SPIE, November 1998.

[Kutter et al, 1999-1] M. Kutter and F.A.P. Petitcolas, "A Fair Benchmark for Image Watermarking Systems", Electronic Imaging: Security and Watermarking of Multimedia Contents, Vol. 3657, pp. 226-239, San Jose, CA, USA, January 1999.

[Kutter et al, 1999-2] M. Kutter, "Performance Improvement of Spread Spectrum Based Image Watermarking Schemes Through M-ary Modulation", in Andreas Pfitzmann (Ed.), Proc. of Information Hiding '99, Vol. LNCS 1768, pp. 237-252, Dresden, September 1999.

L

[Langelaar et al, 1996] G.C. Langelaar, J.C.A. van der Lubbe and J. Biemond, "Copy Protection for Multimedia Data Based on Labelling Techniques", Proc. 17th Symposium on Information Theory in the Benelux, Enschede, The Netherlands, May 1996.

[Langelaar et al, 1997] G.C. Langelaar, R.L. Lagendijk and J.C.A. van der Lubbe, "Robust Labelling Methods for Copy Protection of Images", In Proceedings of SPIE, Conference on Storage and Retrieval for Image and Video Databases V, Vol. 3022, pp. 298-309, San Jose, CA, February 1997.

[Langelaar et al, 2000] G.C. Langelaar, I. Setyawan and R.L. Lagendijk, "Watermarking Digital Image and Video Data: A state-of-the-art Overview", IEEE Signal Processing Magazine, Vol. 17, pp. 20-46, September 2000.

[Lee et al, 2000] C.H. Lee, H.S. Oh and H.K. Lee, "Adaptive Video Watermarking Using Motion Information", Proceedings of SPIE, Electronic Imaging 2000, Vol. 3972, San Jose, CA, USA, January 2000.

[Legge et al, 1980] Gordon E. Legge and John M. Foley, "Contrast Masking in Human Vision", In Journal of the Optical Society of America, Vol. 70, No. 12, pp. 1458-1471, December 1980.

[Lewis et al, 1992] A.S. Lewis and G. Knowles, "Image Compression Using the 2-D Wavelet Transform", In IEEE Transactions on Image Processing, Vol. 1, No. 2, pp. 244-250, April 1992.

[Liang et al, 2000] Te-Shen Liang and Jeffrey J. Rodríguez, "Improved Watermark Robustness Via Spectrum Equalization", In IEEE Intl. Conf. on Acoustics, Speech and Signal Processing ICASSP'2000, Vol. 4, pp. 1951-1954, Istanbul, Turkey, 5-9 June, 2000.

[Licks, 1999] V. Licks, "Image Watermarking in the Fourier Domain", Digital Image Processing Project Report, EECE 533, University of New Mexico, 17 December, 1999.

[Licks et al, 2000] V. Licks and R. Jordan, "On Digital Image Watermarking Robust to Geometric Transformations", accepted for IEEE International Conference on Image Processing 2000, Vancouver, Canada, 2000.

[Lin et al, 1998] C.Y. Lin and S.F. Chang, "A Watermark-Based Robust Image Authentication Method Using Wavelets", In ADVENT Project Report, Columbia University, USA, April 1998.

[Lin et al, 2000] C-Y. Lin, M. Wu, J.A. Bloom, M.L. Miller, I.J. Cox and Y-M. Lui, "Rotation, Scale, and Translation Resilient Public Watermarking for Images", In Proc. SPIE Security and Watermarking of Multimedia Contents II, SPIE EI, 2000.

[Linnartz et al, 1997] J.P.M.G. Linnartz, A.C.C. Kalker, G.F. Depovere and R. Beuker. "A Reliability Model for Detection of Electronic Watermarks in Digital Images", In Proc. Benelux Symposium on Communication Theory, pp. 202-208, Enschede, October 1997.

[Loo et al, 2000] P. Loo and N.G. Kingsbury, "Digital Watermarking With Complex Wavelets", Proc. IEE Colloquium on Secure Images and Image Authentication, IEE, London, 10 April, 2000.

[Lu et al, 1999] C. S. Lu, H.Y. Mark Liao, S. K. Huang and C.J. Sze, "Highly Robust Image Watermarking Using Complementary Modulations", Proc. 2'nd Int. Information Security Workshop, Malaysia, Lecture Notes in Computer Science, Vol. 1729, pp. 136-153, 1999.

[Lu et al, 2000] C.S. Lu, S.K. Huang, C.J. Sze and H.Y.M. Liao, "Cocktail Watermarking for Digital Image Protection", In IEEE Trans. on Multimedia, Vol. 2, No. 4, pp. 209-224, 2000.

[Lumini et al, 2000] A. Lumini and D. Maio, "Blind Watermarking System for Digital Images in the Wavelet Domain", In Proc. 12'th International Symposium Electronic Imaging Security and Watermarking of Multimedia Contents II (EI'00), pp.524-535, San Jose, CA, January 2000.

M

[Maes et al, 2000] M. Maes, T. Kalker, J.P. Linnartz, J. Talstra, G.F.G. Depovere and J. Haitsma, "Digital Watermarking for DVD Video Copy Protection", IEEE Signal Processing Magazine, Vol. 17, No. 5, pp. 47-57, 2000.

[Matsui et al, 1994] Kineo Matsui and Kiyoshi Tanaka, "Video-steganography: How to Secretly Embed a Signature in a Picture", Journal of the Interactive Multimedia Association Intellectual Property Project, Vol. 1, No. 1, pp. 187-205, January 1994.

[Meerwald, 2002] Peter Meerwald, "Digital Watermarking World" Forum, Public discussion, http://www.watermarkingworld.org/

[Misiti et al, 2001] M. Misiti, Y. Misiti, G. Oppenheim, and J-M. Poggi, Matlab™ - The Wavelet Toolbox Manual, Version 2, MathWorks Inc., 2001.

[Mintzer et al, 1998] Fred Mintzer, Gordon W. Braudaway and Alan E. Bell, "Opportunities for Watermark Standards", Communications of the ACM, Vol. 41, pp. 57-64, July 1998.

[Mintzer et al, 1999] F. Mintzer and G. Braudaway, "If One Watermark is Good, Are More Better?", Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Vol. 4, Phoenix, Arizona, May 1999.

[Mittelholzer, 1999] Thomas Mittelholzer, "An Information-Theoretic Approach to Steganography and Watermarking", Proceedings of 3'rd International Workshop on Information Hiding IH'99, Andreas Pfitzmann Editor, pp. 1-16, Dresden, Germany, 29 September - 1 October, 1999, LNCS, Vol. 1768, Springer-Verlag, 2000.

[Moulin, 2001] P. Moulin, "The Role of Information Theory in Watermarking and Its Application to Image Watermarking", Signal Processing, Special Issue on Information-Theoretic Issues in Digital Watermarking, Vol. 81, June 2001.

N

[Nikolaidis et al, 1996] Nikos Nikolaidis and Ioannis Pitas, "Copyright Protection of Images Using Robust Digital Signatures", In IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP'96, Vol. 4, pp. 2168-2171, May 1996.

[Nikolaidis et al, 1998] N. Nikolaidis and I. Pitas, "Robust Image Watermarking in the Spatial Domain", Signal Processing, Vol. 66, No. 3, pp. 385-403, May 1998.

[Nickl et al, 1997] H. Nickl, J. Hagenauer and F. Burkert, "Approaching Shannon's Capacity Limit by 0.27dB Using Simple Hamming Codes", IEEE Communication Letters, Vol. 1, pp. 130-132, September 1997.

O

[O'Ruanaidh et al, 1996] J. O'Ruanaidh, W. Dowling and F. Boland, "Phase Watermarking of Digital Images", Proc. IEEE, Int. Conference on Image Processing ICIP'96, Vol. III, pp. 239-242, Lausanne, Switzerland, 16-19 September, 1996.

[O'Ruanaidh et al, 1998] Joseph J.K. O'Ruanaidh and Thierry Pun, "Rotation, Scale and Translation Invariant Spread Spectrum Digital Image Watermarking", Signal Processing, Vol. 66, No. 3, pp. 303-317, May 1998.

P

[Pech-Pacheco et al, 199x] J.L. Pech-Pacheco, G. Cristobal, J. Alvarez-Borrego and M. Keil, "Automatic Object Identification Irrespective to Geometric Changes", Submitted to Applied Optics.

[Pereira et al, 1999] S. Pereira, J.J.K. O'Ruanaidh, F. Deguillaume, G. Csurka and T. Pun, "Template Based Recovery of Fourier-based Watermarks Using Log-polar and Log-log Maps", IEEE Int. Conf. on Multimedia Computing and Systems, Special Session on Multimedia Data Security and Watermarking, Florence, Italy, June 1999.

[Pereira et al, 2000] Shelby Pereira, Sviatoslav Voloshynovskiy and Thierry Pun, "Optimized Wavelet Domain Watermark Embedding Strategy Using Linear Programming", In Harold H. Szu and Martin Vetterli Editors, Wavelet Applications VII (part of SPIE AeroSense 2000), Orlando, Florida, USA, 26-28 April, 2000.

[Pereira et al, 2001] S. Pereira, S. Voloshynovskiy and T. Pun, "Optimal Transform Domain Watermark Embedding Via Linear Programming", Signal Processing, Special Section on Information Theoretic Aspects of Digital Watermarking, 2001. V. Cappellini, M. Barni, F. Bartolini, Eds., Vol. 81, No. 6, pp. 1251-1260, June 2001.

[Perez-Gonzales et al, 1999] Fernando Pérez-González and Juan R. Hernández, "A Tutorial on Digital Watermarking", In Proc. of the 33rd IEEE Annual Carnahan Conference on Security Technology, Madrid, Spain, October 1999.

[Perez-Gonzales et al, 2001] Fernando Pérez-González, Juan R. Hernández and Félix Balado, "Approaching the Capacity Limit in Image Watermarking: A Perspective on Coding Techniques for Data Hiding Applications", Signal Processing, Special Section on Information Theoretic Aspects of Digital Watermarking, Vol. 81, No. 6, pp. 1215-1238, June 2001.

[Peterson et al, 1993] Heidi Peterson, Albert J. Ahumada and Andrew B. Watson, "An Improved Detection Model for DCT Coefficient Quantization" In Human Vision, Visual Processing, and Digital Display IV, The International Society for Optical Engineering, SPIE, Vol. 1913, pp. 191-201, 1993.

[Petitcolas et al, 1998] F. Petitcolas and R. Anderson, "Weaknesses of copyright marking systems", Proceedings of the ACM Multimedia and Security Workshop (at ACM Multimedia '98), Dittmann et al, Editors, pp. 55-62, Bristol, United Kingdom, September 1998.

[Petitcolas et al, 1999-1] F.A.P. Petitcolas and R.J. Anderson, "Evaluation of copyright marking systems", In International Conference on Multimedia Systems ICMS'99, pp. 574-579, Florence, Italy, 7-11 June, 1999.

[Petitcolas et al, 1999-2] Fabien A.P. Petitcolas, Ross J. Anderson and Markus J. Kuhn, "Information Hiding: A Survey", Proc. of the IEEE, Vol. 87, No. 7, pp. 1062-1078, July 1999.

[Pitas et al, 1995] I. Pitas and T.H. Kaskalis, "Applying Signatures on Digital Images", Proceedings of IEEE Workshop on Nonlinear Signal and Image Processing, pp. 460-463, Neos Marmaras, Greece, 20-22 June 1995.

[Pitas, 1996] I. Pitas, "A Method for Signature Casting on Digital Images", In Proc. of the IEEE Int. Conf. on Image Processing, pp. 215-218, Lausanne, Switzerland, September 1996.

[Piva et al, 1998] A. Piva, M. Barni, F. Bartolini and V. Cappellini, "Threshold Selection for Correlation-based Watermark Detection", Proceedings of COST 254 Workshop on Intelligent Communications, pp. 67-72, L'Aquila, Italy, 4-6 June, 1998.

[Piva et al, 2000] A. Piva, M. Barni, F. Bartolini, V. Cappellini and A. De Rosa, "Improving the Robustness of Non-additive Watermarks Through Optimum Detection Theory", In Security and Watermarking of Multimedia Contents II, Wong, Delp, Editors, Proceedings of SPIE, Vol. 3971, pp. 2-13, San Jose, CA, January 2000.

[Podilchuk et al, 1997-1] C.I. Podilchuk and W. Zeng, "Perceptual Watermarking of Still Images", Proc. of the 1'st IEEE Signal Processing Society Workshop on Multimedia Signal Processing, Princeton, New Jersey, USA, June 1997.

[Podilchuk et al, 1997-2] C.I. Podilchuk and W. Zeng, "Digital Image Watermarking Using Visual Models", In Human Vision and Electronic Imaging II, B.E. Rogowitz and T.N. Pappas Editors, IS&T and SPIE, Vol. 3016, pp. 100-111, San Jose, CA, 1997.

[Podilchuk et al, 1998] C.I. Podilchuk and W. Zeng, "Image-Adaptive Watermarking Using Visual Models", IEEE Transactions on Selected Areas of Communications, Vol. 16, No. 4, pp. 525-539, May 1998.

[Press et al, 1992] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, "Numerical Recipes in C: The Art of Scientific Computing, Second Edition, Cambridge University Press, UK, 1992.

Q

[Queluz et al, 2000] M.P. Queluz and P. Lamy, "Spatial Watermark for Image Verification", IST&SPIE - Electronic Imaging 2000, San Jose, January 2000.

R

[Ramkumar et al, 1998-1] M. Ramkumar, A.N. Akansu, "A Robust Scheme for Oblivious Detection of Watermarks / Data Hiding in Still Images", SPIE Multimedia Systems and Applications, Boston, MA, Vol. 3528, pp. 474-481, November 1998.

[Ramkumar et al, 1998-2] M. Ramkumar and A.N. Akansu, "Theoretical Capacity Measures for Data Hiding in Compressed Images", In Proc. SPIE, Voice, Video and Data Communications, Vol. 3528, pp. 482-492, November 1998.

[Ramkumar, 1999] M. Ramkumar, "Data Hiding in Multimedia: Theory and Applications", PhD thesis, New Jersey Institute of Technology, Kearny, NJ, USA, November 1999.

[Ramkumar et al, 1999] M. Ramkumar, A.N. Akansu and A.A. Alatan, "A Robust Data Hiding Scheme for Images Using DFT", In Proceedings of the 6'th IEEE International Conference on Image Processing ICIP '99, pp. 211-215, Kobe, Japan, October 1999.

[Reddy et al, 1996] B.S. Reddy and B.N. Chatterji, "An FFT-based Technique for Translation, Rotation and Scale-invariant Image Registration", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 5, No. 8, pp. 1266-1270, August 1996.

[Robert et al, 2000] Arnaud Robert and Raymond Knopp, "Detection Theory and Digital Watermarking", In Proc. SPIE, Security and Watermarking of Multimedia Contents II, Vol. 3971, pp. 14-23, 24-26 January, 2000.

S

[Schneier, 1996] Bruce Schneier, "Applied Cryptography: Protocols, Algorithms and Source Code in C", Second Edition, John Wiley and Sons Inc., New York, USA, 1996.

[Schyndel et al, 1994] R.G. van Schyndel, A.Z. Tirkel, N.R.A. Mee and C.F. Osborne, "A Digital Watermark", In Proceedings of the IEEE International Conference on Image Processing, Vol. 2, pp. 86-90, Austin, Texas, USA, November 1994.

[Singh, 1999] Simon Singh, "The Code Book: The Secret History of Codes and Code-breaking", Fourth Estate, London, 1999.

[Smith et al, 1996] J.R. Smith and B.O. Comisky, "Modulation and Information Hiding in Images", In Proc. 1'st Int. Workshop on Information Hiding, R. Anderson Ed., Cambridge, UK, May/June 1996, LNCS, Vol. 1174, pp. 207-226, Springer-Verlag, Berlin, 1996.

[Solachidis et al, 1999] V. Solachidis and I. Pitas, "Circularly Symmetric Watermark Embedding in 2-D DFT Domain", Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing ICASSP'99, Vol. 6, pp. 3469-3472, Phoenix, Arizona, USA, March 1999.

[Stone, 1996] H.S. Stone, "Analysis of Attacks on Image Watermarks with Randomized Coefficients", NEC Research Institute, Technical Report, 1996.

[Swanson et al, 1996-1] Mitchell D. Swanson, Bin Zhu and Ahmed H. Tewfik, "Robust Data Hiding for Images", In 7'th Digital Signal Processing Workshop DSP'96, pp. 37-40, Loen, Norway, September 1996.

[Swanson et al, 1996-2] M.D. Swanson, B. Zhu and A. Tewfik, "Transparent Robust Image Watermarking", In Proceedings of the IEEE Int. Conf. on Image Processing ICIP'96, pp. 211-214, Lausanne, Switzerland, September 1996.

[Swanson et al, 1997] M.D. Swanson, B. Zhu, B. Chau and A.H. Tewfik, "Object-based Transparent Video Watermarking", IEEE Workshop on Multimedia Signal Processing, pp. 369-374, Princeton, USA, June 1997.

[Swanson et al, 1998-1] Mitchell D. Swanson, Mei Kobayashi and Ahmed H. Tewfik, "Multimedia Data-embedding and Watermarking Technologies", Proceedings of the IEEE, Vol. 86, No. 6, pp. 1064-1087, June 1998.

[Swanson et al, 1998-2] M.D. Swanson, B. Zhu and A.H. Tewfik, "Multiresolution Scene-based Video Watermarking Using Perceptual Models", IEEE Journal of Selected Areas of Communications, Vol. 16, No. 4, pp. 540-550, May 1998.

T

[Tanaka et al, 1990] K. Tanaka, Y. Nakamura and K. Matsui, "Embedding Secret Information Into a Dithered Multilevel Image", In Proceedings of the IEEE Military Communications Conference, pp. 216-220, September 1990.

[Tao et al, 1997] Bo Tao and Bradley Dickinson, "Adaptive Watermarking in the DCT Domain", Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing ICASSP'97, Munich, Germany, 21-24 April, 1997.

[Tsekeridou et al, 2000] S. Tsekeridou and I. Pitas, "Embedding Self-Similar Watermarks in the Wavelet Domain", IEEE Int. Conf. on Acoustics, Systems and Signal Processing ICASSP'2000, Vol. IV, pp. 1967-1970, Istanbul, Turkey, 5-9 June, 2000.

[Tirkel et al, 1993] A.Z. Tirkel, G.A. Rankin, R.M. van Schyndel, W.J. Ho, N.R.A. Mee and C.F. Osborne, "Electronic Watermark", In Digital Image Computing, Technology and Applications DICTA'93, pp. 666-673, Macquarie University, Sidney, Australia, 1993.

V

[Villasenor et al, 1995] John D. Villasenor, Benjamin Belzer and Judy Liao, "Wavelet Filter Evaluation for Image Compression", IEEE Transactions on Image Processing, Vol. 4, No. 8, pp. 1053-1060, August 1995.

W

[Wade, 2000] Graham Wade, "Coding Techniques: An Introduction to Compression and Error Control", Macmillan Publishing, 2000.

[Wang et al, 1998] Houng-Jyh Wang, Po-Chyi Su and C.C. Jay Kuo, "Wavelet-based Digital Image Watermarking", Optics Express, Vol. 3, No. 12, pp.491-496, December 1998.

[Watson, 1993] A.B. Watson, "DCT Quantization Matrices Visually Optimized for Individual Images", Proc. of the SPIE Conference on Human Vision, Visual Processing and Digital Display IV, pp. 202-216, San Jose, CA, February 1993.

[Watson et al, 1994] A.B. Watson, J.A. Solomon and A. Ahumada, "DCT Basis Function Visibility: Effects of Viewing Distance and Contrast Masking", In Proceedings of the SPIE, Vol. 2179, pp. 99-108, 1994.

[Watson et al, 1996] A.B. Watson, G.Y. Yang, J.A. Solomon, and J. Villasenor, "Visual Thresholds for Wavelet Quantization Error", In Proceedings of the SPIE, Human Vision and Electronic Imaging, B. Rogowitz and J. Allebach Editors, The Society for Imaging Science and Technology, Vol. 2657, pp. 382-392, 1996.

[Watson et al, 1997] A.B. Watson, G.Y. Yang, J.A. Solomon and J.D. Villasenor, "Visibility of Wavelet Quantization Noise", IEEE Transactions on Image Processing, Vol. 6, No. 8, pp. 1164-1175, August 1997.

[Wolfgang et al, 1997] R.B. Wolfgang and E.J. Delp, "Overview of Image Security Techniques With Applications in Multimedia Systems", Proceedings of the SPIE Conference on Multimedia Networks, Vol. 3228, pp. 297-308, Dallas, TX, USA, November 1997.

[Wolfgang et al, 1999] R.B. Wolfgang, C.I. Podilchuk and E.J. Delp, "Perceptual Watermarks for Digital Images and Video", Proceedings of the IEEE, Vol. 87, No. 7, pp. 1108-1126, July 1999.

[Wolberg et al, 2000] George Wolberg and Siavash Zokai, "Robust Image Registration Using Log-Polar Transform", Proceedings of the IEEE, International Conference on Image Processing ICIP'2000, Vancouver, Canada, September 2000.

[Wood, 2000] David Wood, "The Challenges of Rights Management", EBU Technical Review, No. 282, March 2000.

[Wozencraft et al, 1965] J.M. Wozencraft and I.M. Jacobs, "Principles of Communication Engineering", John Wiley and Sons, New York, USA, 1965.

X

[Xia et al, 1998] X. Xia, C.G. Boncelet Jr. and G.A. Arce, "Wavelet Transform Based Watermark for Digital Images", Optics Express, Vol. 3, No. 12, December 1998.

[Xie et al, 2000] Hongjie Xie, Nigel Hicks, G. Randy Keller, Haitao Huang, and Vladik Kreinovich, "Automatic Image Registration Based on a FFT Algorithm and IDL/ENVI", Proceedings of the ICORG-2000 International Conference on Remote Sensing and GIS/GPS, Vol. 1, pp. 397-402, Hyderabad, India, 1-4 December, 2000.

Z

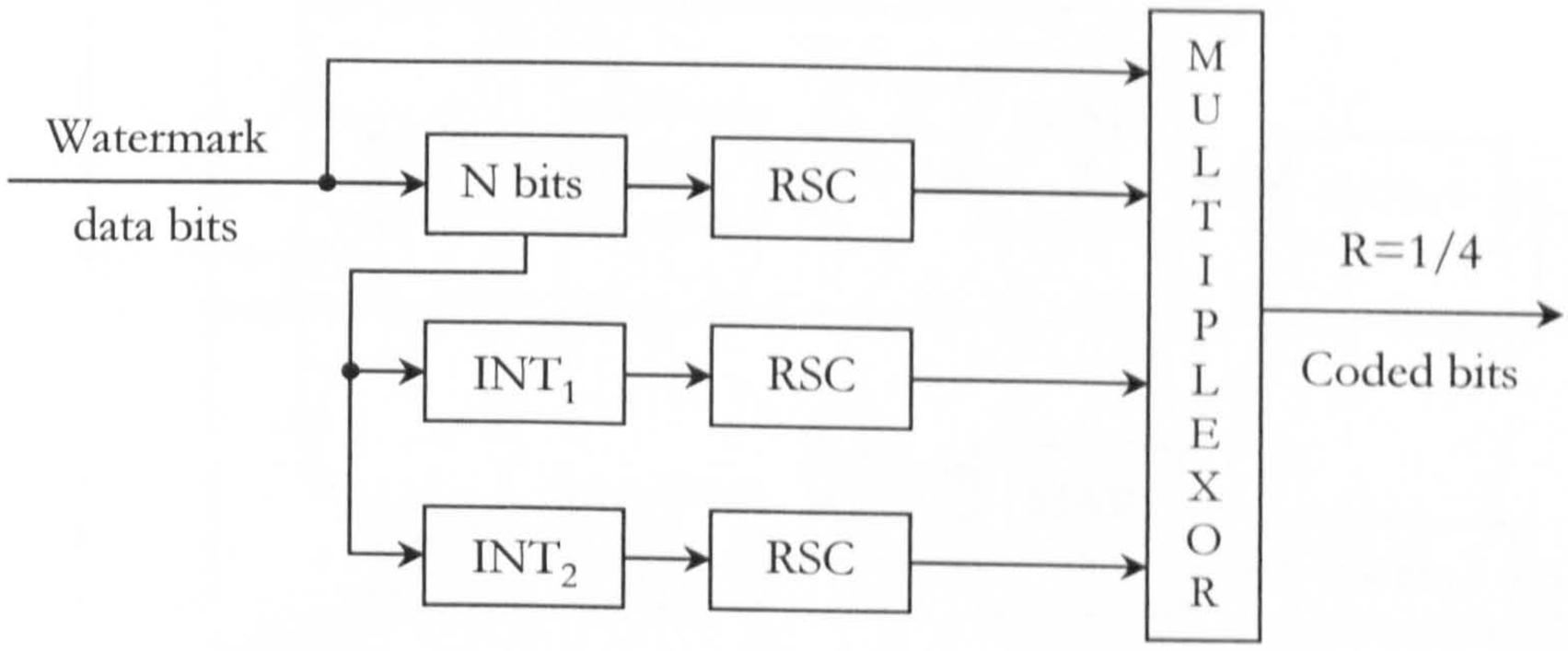
[Zhu et al, 1995] B. Zhu, A.H. Tewfik and O.N. Gerek, "Low Bitrate Near-Transparent Image Coding", In Proc. of the SPIE Int. Conf. on Wavelet Applications for Dual Use, Vol. 2491, pp. 173-184, Orlando, FL, 1995.

[Zhu et al, 1996] B. Zhu, M.D. Swanson and A.H. Tewfik, "A Transparent Robust Authentication and Distortion Measurement Technique for Images", In Proc. IEEE Digital Signal Processing Workshop, pp. 45-48, Loen, Norway, September 1996.

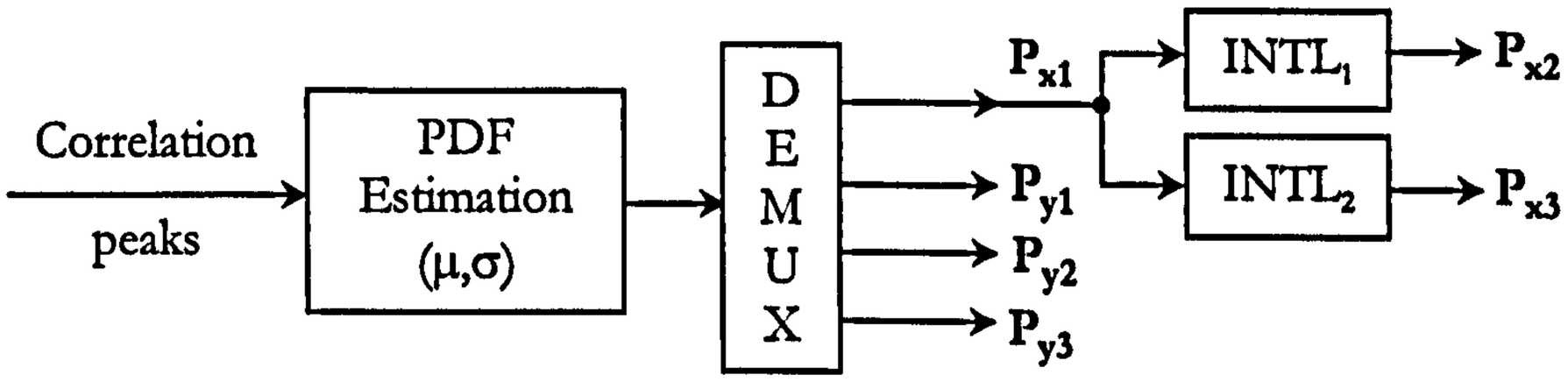
Appendix A

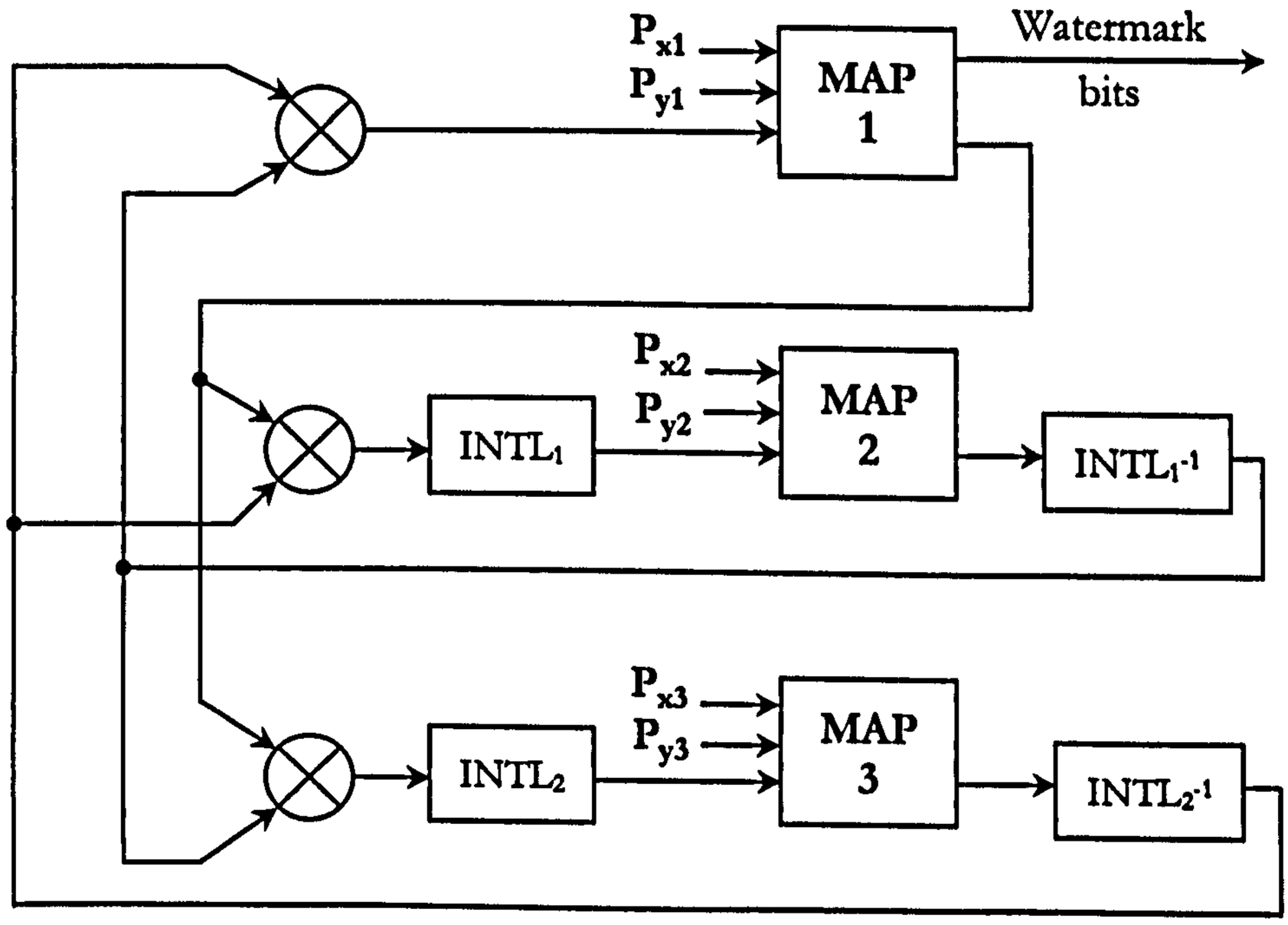
The 3PCCC Turbo Code

A.1 The Turbo Encoder



A.2 The Turbo Decoder





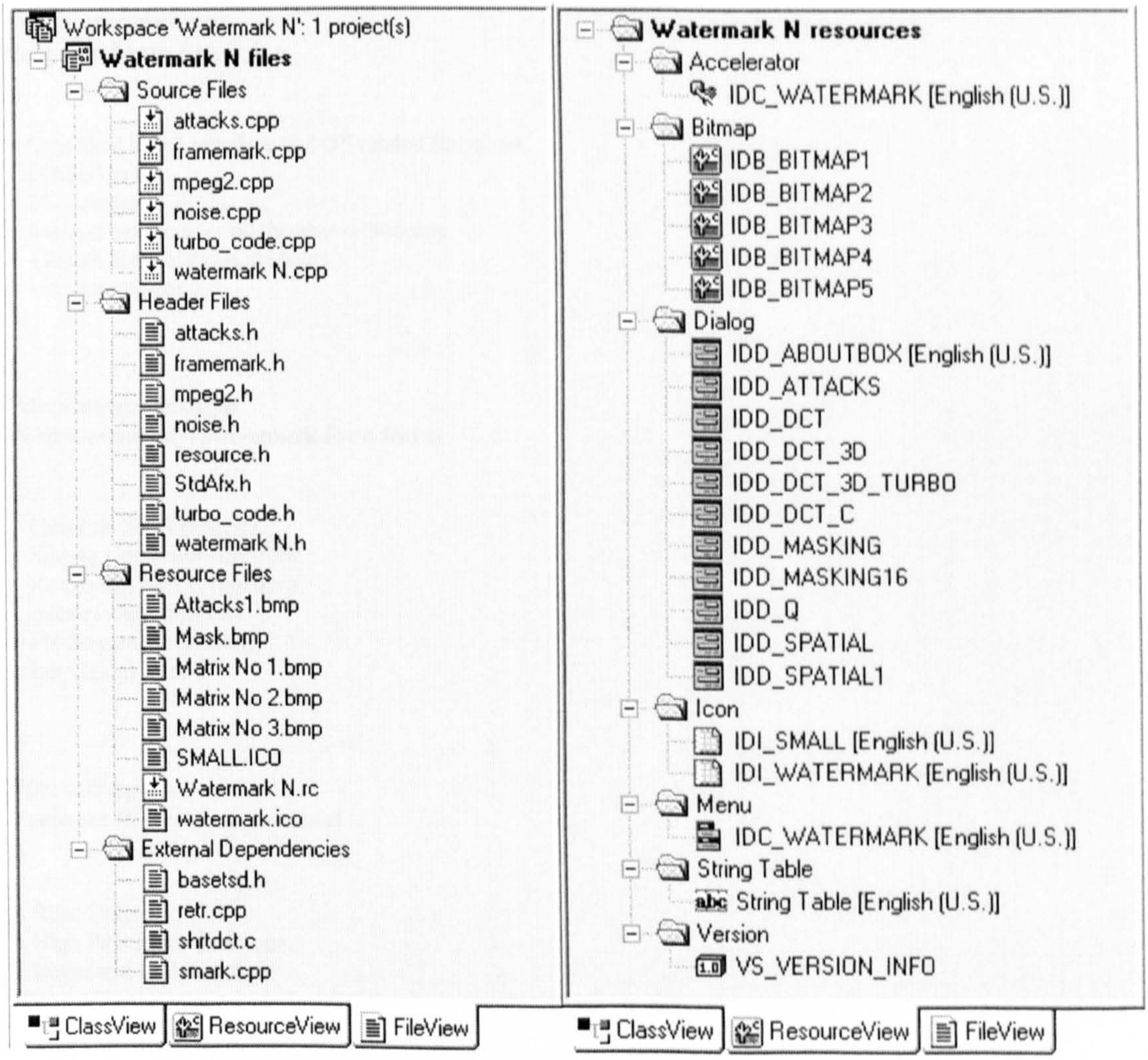
Appendix

B

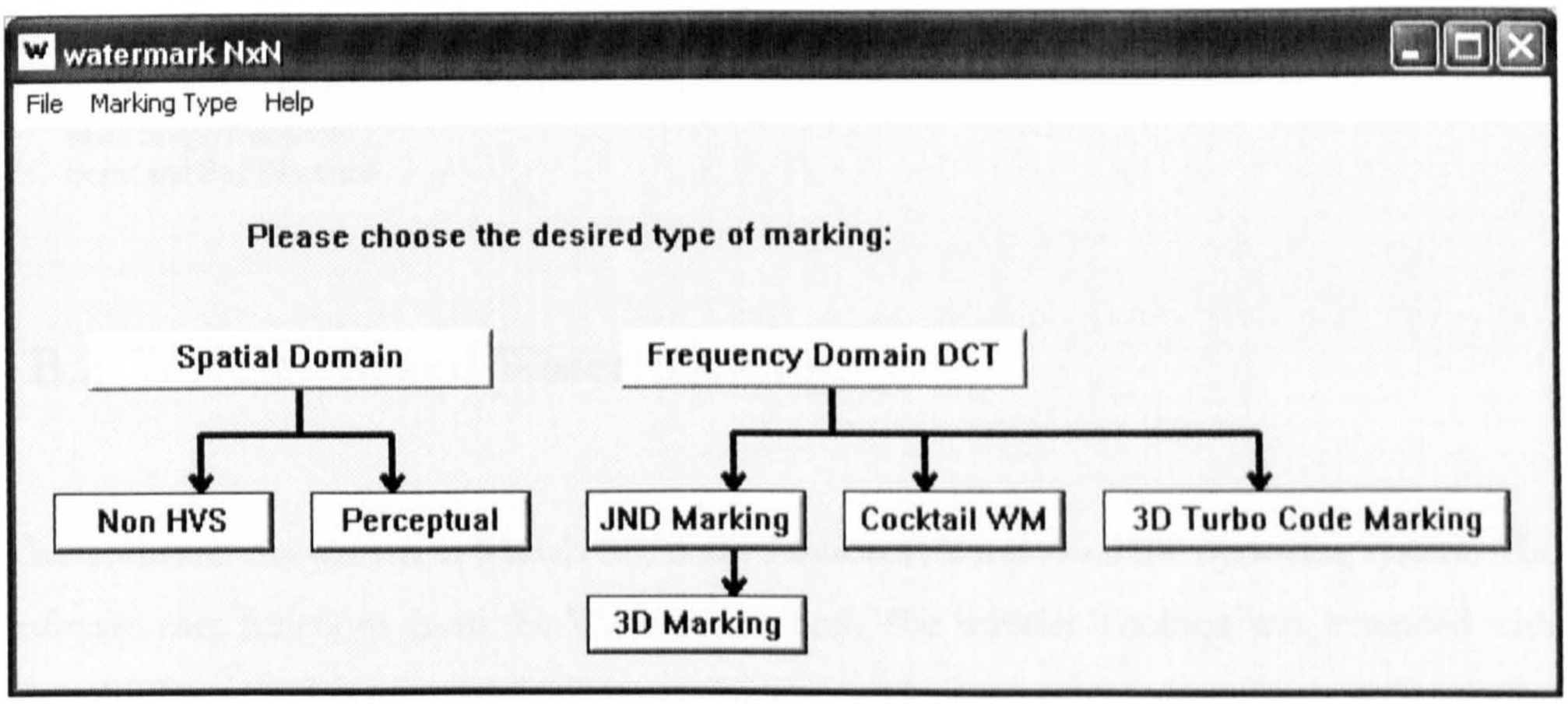
Software

B.1 Spatial Domain & DCT-Based Watermarking

This software was written in Microsoft Visual C++ 6.0 (SP 5), uses Win32 API functions and therefore it is intended to work under Microsoft 98, ME, NT4, 2000, XP operating systems only. A list of modules and resources used by the program is provided below:



The program has a modular architecture, as illustrated below:



The summary of the main component files and their content is briefly described in the following paragraphs. The main body of the program is located in the Watermark N.cpp file.

```
/*
 * File: watermark.cpp
 * Purpose: Main file.
*/

//    Graphical (user) interface and OS related functions.
//    Global vars.
//    I/O functions.
//    Mesage handlers for all the above modules.
//    Threads for the above modules.
//    Various utilities.


/*
 * File: framemark.cpp
 * Purpose: supply watermark for a frame
*/

//    Generate SS Watermark
//    Sliding Correlator functions
//    Retrieving Report functions
//    Interleaving functions
//    PN Sequence functions
//    Init Global Vars


/*
 * File: retr.cpp
 * Purpose: Watermark retrieval
*/

//    Read Frame
//    High Pass Filter functions
//    Begin and End functions
```

```
/*
 * File: smark.cpp
 * Purpose: Watermark retrieval
*/

//    Mark Image functions
//    Begin and End functions
```

B.2 Wavelet-Based Watermarking

This software was written in Matlab 6.0, under Microsoft Windows 2000 operating system. The software uses functions from the Wavelet Toolbox. The wavelet Toolbox was extended with several other wavelets (Antonini 7.9 for example).
A list of the main functions and their role is provided below:

```
function out = add_mark(in, aorc, lel, orient, dd, q, alpha, data, dlen)
%
%  out = in * alpha * HVS(q) * databit * PN
%
%  Usage:  out = add_mark(in, aorc, lel, orient, q, alpha, data, dlen)
%
%  The parameters are:     out, in -> no comment needed
%                          aorc -> if zero, means that we are processing appcoef
%                                  else, we are processing detcoef
%                          lel -> current level, between 1 ... N
%                          orient -> orientation, in case of aorc~=0
%                          d -> the crt number of the databit
%                          q -> the Q matrix from the Watson's HVS model; ft(lel, orient)
%                          alpha -> strength adjusting for WM; scalling factor for q()
%                          data -> the WM data bits
%                          dlen -> the nr. Of WM data bits


function DWTMarking(method, level, alpha, datalen)
%
%  Usage: DWTMarking(method, level, alpha, datalen);
%                  method = 1 <---> Watson, an7.9, level=[2..4]
%                  level = 2..4
%                  alpha = the strength of marking
%                  datalen = the number of input data bits


function DWTRetrieving(method, level, datalen, attack, a_param, b_param)
%
%  Usage: DWTRetrieving(method, level, datalength, attack, a_param, b_param);
%                  method = 1 <---> Watson, an7.9, level=[2..4]
%                  level = 2..4
%                  datalength – no comment
%                  attack = 1 – JPEG compression:
%                                  a_param = the quality factor (1..100), b_param = 0;
%                           = 2 – Scalling and rescalling back:
%                                  a_param = the scalling factor (+/- 1/x)
%                                  b_param = the scalling method 'nearest', 'bilinear', 'bicubic'
%                           = 3 – Scalling with a small % without rescalling back:
%                                  a_param = idem 2 (e.g +/- 1/100)
%                                  b_param = idem 2
```

```
%                   = 4 – Cropping:
%                       a_param = [xmin ymin width height];
%                       b_param = 0;
%                   = 5 – Shifting:
%                       a_param = H Shift;
%                       b_param = V Shift;


function out1 = DWTXcorr(in, d, aorc, N, lel, orient, q, dlen)
function BER = EbNo2BER(LUT, EbNo)
function [out, out1] = FFTXcorr(in, pnmat, xcorr_type)
function [out, out1] = FFTXcorrNew(in, pnmat)
%
%  Various supporting functions & utilities
%

function map = llm(img,res)
%
%  Cartesian -> Log Log conversion, Greyscale interpolation used
%  img = Greyscale image, as given from imread for example
%  res = resolution
%  usage: map = llm(image,600);


function lpv = logpolar(img)
%
%  Cartesian -> Log Polar conversion, Greyscale interpolation used
%  img = Greyscale image
%  usage: lpv = logpolar(image);


function out_img = sprefmark(in_img, pnmat);
%
%  Insert a spatial watermark as a reference
%


function testdwtxcorr(start,frame,filename)
function out = testMPEG2Reg(start,frame,filename,bit_rate)
function out = testMPEG2Wav(start,frame,filename,bit_rate,datalen)
file TestXcorrShifts.m
%
%  Main functions; Batch simulation, various attacks
%


function q = watsonDWT(display_rez, viewing_dist, level)
%
%   USAGE: q = watsonDWT(display_rez, viewing_dist, level);
%
%         q(level, orientation) → level = {1...6}
%                                 orientation = {LL, HL, HH, LH}
%
%                                 |-------|-------|
%                                 | 1 LL  | HL 2  |
%                                 |-------|-------|
%                                 | 4 LH  | HH 3  |
%                                 |-------|-------|
%
%   !!! THE WAVELET TRANSFORM MUST BE ANTONINI 9/7 !!!
%


function marked_image = WaveMarkWatson(image, level, alpha, datalen)
%
%  Watermarking an image/frame in wavelet domain
%
```

THE ENTIRE C++ AND MATLAB CODE IS PROVIDED ON THE ATTACHED CD.

Copyright, License and Disclaimer

This program/code is FREE SOFTWARE; it can be redistributed and/or modified under the terms of the GNU General Public License as published by the Free Software Foundation.

This program/code is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY. Please see the GNU General Public License for more details. A copy of the GNU General Public License can be found on the attached CD.

This code is NOT Public Domain software. The software can be freely distributed, but THE AUTOR RETAINS OWNERSHIP AND COPYRIGHT OF THE SOFTWARE AND ITS SOURCE CODE IN ITS ENTIRETY.

This software can be freely used and/or distributed ONLY FOR NON-PROFIT PURPOSES (educational); ANY COMMERCIAL USE of this software/code or even the use of parts of this software in a commercial product is STRICTLY PROHIBITED without the prior WRITTEN consent of the author.

Appendix C

Publications

C.1 "Adding Robustness to Geometrical Attacks to a Wavelet Based, Blind Video watermarking System"

C.2 "Protecting Intellectual Rights: Digital WM in the Wavelet Domain"

C.3 "DWT Based High Capacity Blind Video Watermarking, Invariant to Geometrical Attacks"

C.4 "Watermarking Uncompressed Video: An Overview"

ADDING ROBUSTNESS TO GEOMETRICAL ATTACKS TO A WAVELET BASED, BLIND VIDEO WATERMARKING SYSTEM

C.V. Serdean, M.A. Ambroze*, M. Tomlinson* and J.G. Wade***

*Department of Communication & Electronic Engineering, University of Plymouth, Plymouth, PL4 8AA, UK

**Department of Electrical & Computer Engineering, University of Newcastle, Callaghan, NSW 2308, Australia

ABSTRACT

This paper describes a high capacity blind video watermarking system invariant to geometrical attacks such as shift, rotation, scaling and cropping. A spatial domain reference watermark is used to obtain invariance to geometric attacks by employing image registration techniques to determine and invert the attacks. A second, high capacity watermark, which carries the data payload, is embedded in the wavelet domain according to a human visual system (HVS) model. This is protected by a state-of-the-art error correction code (Turbo code). For a false detection probability of 10^{-8}, the proposed system is invariant to scaling up to 180%, rotation up to 70^0, and arbitrary aspect ratio changes up to 200% on both axes. Furthermore, the system is virtually invariant to any shifting, cropping, or combined shifting and cropping attack, and it is robust to MPEG2 compression as low as 2-3Mbps.

1. INTRODUCTION

One of the most difficult problems in digital video watermarking is watermark recovery in the presence of geometric attacks like frame shift, cropping, scaling, rotation, and change of aspect ratio, especially when some of these are combined together. The work presented in this paper was carried out in the context of uncompressed video (ITU-R 601 as found in TV studios), and so geometric attacks tend to be less severe compared to those for image watermarking [1]. On the other hand, the recovery problem is compounded for video since it must be carried out blind due to the difficulty of storing the original. In this case, for the typical spread spectrum (SS) watermarking system, blind retrieval is performed via cross-correlation between the marked video and the secret pseudo-noise (PN) sequence used to spread the watermark at the embedding stage. Recovery is straightforward given perfect synchronisation between the attacked video and the PN sequence, but is difficult when geometric attacks destroy the synchronisation. In this case it is possible to perform some form of sliding correlation in order to re-establish synchronisation i.e. multiple cross-correlations over a specified search space. Unfortunately the search space grows very quickly, making it difficult to recover the watermark in a reasonable time. Clearly, given that retrieval in a video context must be done in near real time, the computational problem is very significant in the presence of attacks. One way to overcome this is to use image registration techniques for resynchronisation.

The Fourier-Mellin transform (FMT) was first used by O'Ruanaidh [3] to achieve rotation, scaling and translation invariance for image watermarking. Unfortunately, marking in the FMT domain has two major drawbacks: the need to compute the inverse log-polar transform (a lossy operation that drastically reduces system performance), and the need to maintain the FFT symmetry, which halves the watermark capacity. An improved technique was later proposed by Lin [4].

Unlike the previous methods, this paper combines the advantages of an algorithm based on the FM transform image registration techniques [2], with watermarking in the Discrete Wavelet Transform (DWT) domain. The idea is to first undo geometric attacks using the FMT approach and an additional spatial reference watermark used only for registration purposes. Once the attack parameters are determined, the geometric attacks are undone and the resulting frame is passed to the main watermark decoder where the embedded data bits are recovered.

The main watermark, which carries multi-bit data, is inserted in the DWT domain and capacity is maximised by embedding based on a HVS model. The complete system can be regarded as a noisy communications channel and so is protected by Turbo coding. The net result is a relatively simple system that can withstand severe geometric attack, the limiting attack being defined by a threshold yielding a false detection probability of 10^{-8}, and capacity being defined by a BER of 10^{-8}.

2. COMBATING GEOMETRIC ATTACK USING LOG-POLAR AND LOG-LOG TRANSFORMATIONS

When registering two images, the noise is relatively small, and so the correlator usually performs very well. The problem is more difficult for video watermarking since the original video frame is not available. To overcome this, we use a SS reference watermark. The PN sequence used to embed this watermark corresponds to the "original image" and the watermark embedded in the unsynchronised marked video corresponds to a noisy "attacked image" (the video itself represents the noise).

In the proposed system we embed two different watermarks. The first is a 1-bit watermark used only for geometric reference, and is embedded in the spatial domain. The

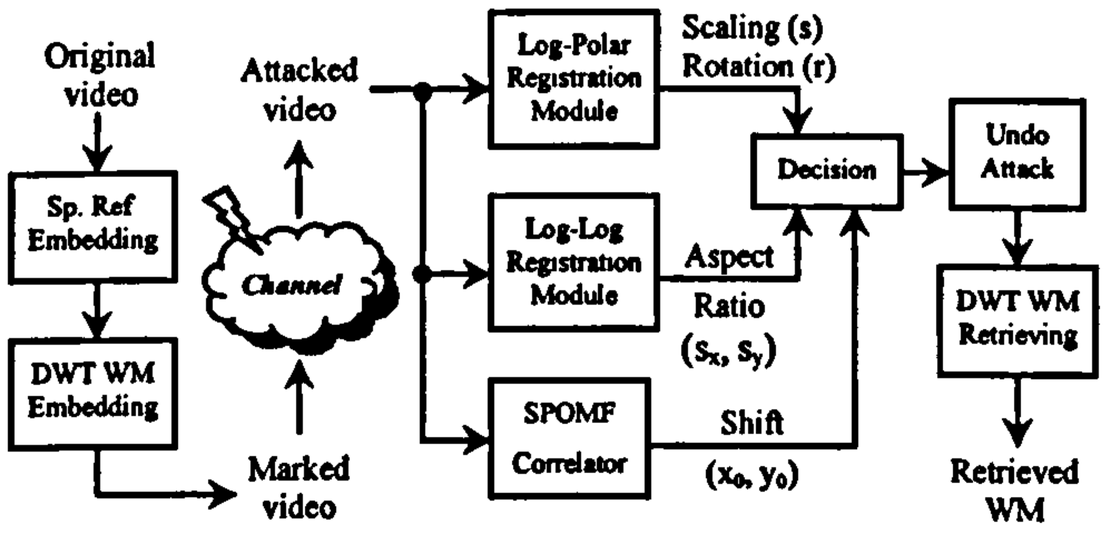


Figure 1. Block schematic of the geometric invariant system

second, multi-bit watermark is used for the data payload, and is embedded in the DWT domain.

The desired geometric invariance can be achieved by using the FMT to convert rotation and scale to spatial shifts [2], which are then easily recovered by a Symmetrical Phase Only Matched Filter (SPOMF).

If frame f_2 is the scaled and rotated replica of frame f_1 with a scaling factor a and angle θ_0, then:

$$f_2(x,y) = f_1\left(a(x\cos\theta_0 + y\sin\theta_0), a(-x\sin\theta_0 + y\cos\theta_0)\right)$$

$$F_2(u,v) = \frac{1}{a^2} F_1\left(\frac{(u\cos\theta_0 + v\sin\theta_0)}{a}, \frac{(-u\sin\theta_0 + v\cos\theta_0)}{a}\right) \quad (1)$$

In order to convert both scaling and rotation to shifts, it is necessary to convert the Cartesian coordinates into log-polar coordinates:

$$x = e^{\log\rho}\cos\theta; \quad y = e^{\log\rho}\sin\theta \quad (2)$$

The result is:

$$F_2(\log\rho, \theta) = F_1(\log\rho - \log a, \theta - \theta_0) \quad (3)$$

where the scale and rotation factors can be retrieved by SPOMF correlation.

Since the FMT is not shift invariant, it is necessary to apply the Fourier magnitude of the frame (rather than the frame itself) to the input of the log-polar conversion module. The Fourier magnitude is shift invariant and so the rotation and scaling parameters can be found even in the presence of shift. After undoing rotation and scaling, the shift is then recovered by performing a simple SPOMF correlation. This technique works well for image-image registration, since the correlation peaks are relatively large and the phase loss can be tolerated. Unfortunately, for video watermarking, the loss can make cross-correlation unreliable, and this approach cannot be used for retrieval under combined attack.

A log-polar map permits recovery over a wide range of scale changes, rotation, or even combined scale-rotation attack. If a log-log map is used, then it is possible to recover arbitrary aspect ratio changes (different scale factors for x and y axes). The shifts alone are easily recovered using a SPOMF module. However, shift recovery from a combined attack (e.g. shift + scaling) requires a comprehensive search for all of the possible shifts, and is computationally intensive.

Fig. 1 shows a schematic of the proposed system. The decision block determines if the reference watermark is present (to within a desired false detection probability), and if present it automatically determines the attack parameters. In the proposed scheme, the two watermarks are embedded in different domains in order to minimise crosstalk, and each watermark is embedded at the full strength dictated by its own HVS model. The reference watermark is embedded in the spatial domain using SS, together with a simple visual model that inserts a stronger watermark in those regions where it is less easily observed (at edges and in high texture regions). The same reference watermark is embedded in all the frames in order to increase the SNR at the correlator input via frame averaging. As a result, the registration takes place only once, and not for each separate frame. This is possible because attacks must be identical for each frame in order to avoid temporal artefacts.

3. THE DWT VIDEO WATERMARKING SCHEME

The hierarchical property of the DWT offers the possibility of analysing a signal at different resolutions (levels) and orientations. This multiresolution analysis gives both space and frequency localisation, and different orientations extract different features of the frame, such as vertical, horizontal, and diagonal information. Generally speaking, edges and textures will be represented by large coefficients in the high frequency sub-bands, and they are well localised within the sub-band. In practice, wavelet analysis is performed using multilevel filter banks. Essentially this comprises a succession of filtering and sub-sampling operations and has been widely described in the literature. For watermarking, we selected the Antonini 7.9 wavelet, as being one of the best wavelets available for image compression [5, 6]. Watermarking in the DWT domain has many advantages compared with FFT or DCT marking [7]. In particular, the multiresolution property provides both local and global spatial support, it is compatible with the HVS, there are no blocking artefacts, and it has lower computational cost and better energy compaction properties than the FFT and DCT.

The information capacity of the channel is maximised by embedding the main payload according to an HVS model, and through the use of Turbo coding. The embedding and retrieving of the watermark are shown in Fig. 2, where we use 3 levels of decomposition. The security of such a system relies in the secret watermarking key, K_1, and in order to improve the system's overall security we use an interleaver to provide a random distribution of the data bits within each sub-band. The interleaver uses a separate key, K_2.

The hierarchical nature of the DWT is exploited by inserting a self-contained watermark in each sub-band, i.e. all payload bits are inserted into each sub-band. The watermark is embedded using amplitude modulation:

$$C_i^M = \begin{cases} C_i + \alpha \underbrace{\frac{Q(\lambda,\theta)}{Q_{\min}} \cdot \frac{|C_i|}{mean(|C_i|)}}_{S} \cdot W_i, & \text{(details)} \\ \quad if \quad S > 24, \quad then \quad S = 24. \\ C_i + \alpha \frac{Q(\lambda,\theta)}{2} \cdot \frac{|C_i|}{mean(|C_i|)} \cdot W_i, & \text{(approximation)} \end{cases} \quad (4)$$

where Q_{min} is the minimum value within matrix Q, W_i is the watermark, C_i is the original wavelet coefficient and C_i^M is the marked coefficient. Note that equation (4) incorporates media dependence, essential for robust watermarking. This marks more heavily the high frequency sub-bands and the largest

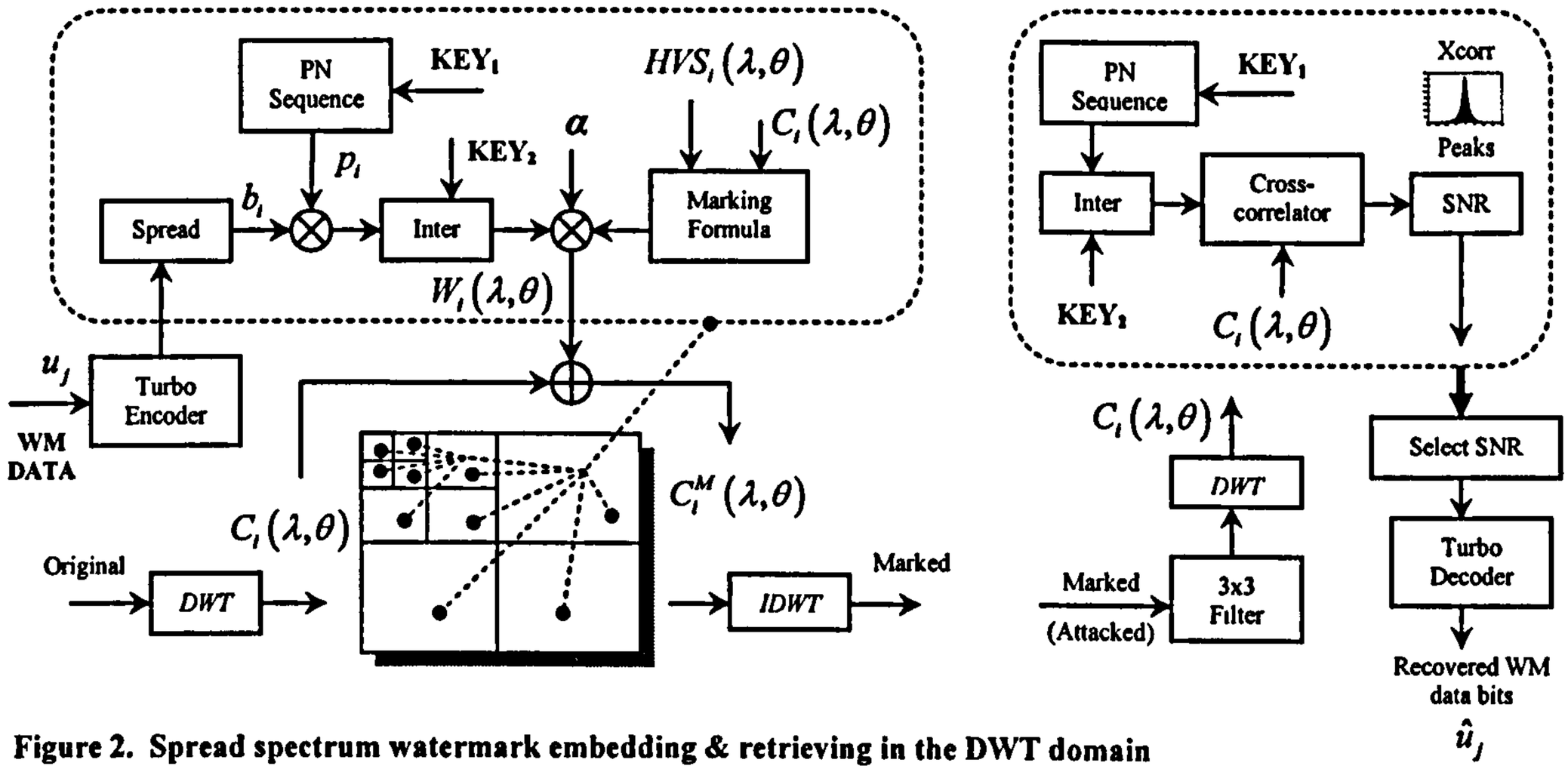


Figure 2. Spread spectrum watermark embedding & retrieving in the DWT domain

coefficients, since modification of these coefficients is less likely to incur visible artefacts. The HVS is incorporated in the quantisation matrix, $Q(\lambda,\theta)$ where λ is the level and θ is the orientation. For computing Q we use a visual model developed by Watson for the Antonini 7.9 DWT [5]. Although this is a much simpler model than those usually used in DCT schemes, the overall performance of the system is actually better.

For retrieval, the video sequence is filtered using a Laplacian 3x3 filter prior to cross-correlation in order to improve the performance of the correlator.

It is advantageous to have a self-contained watermark in each sub-band, since a SNR can be determined for each sub-band as an indicator of sub-channel quality. Different types of attack affect different levels and orientations in different ways, and so it is always possible to select an optimal sub-band via SNR. Correlation is therefore performed separately for each sub-band, obtaining a set of cross-correlation peaks (one peak for each embedded data bit) for each sub-band. A SNR is then computed for each set of cross-correlation peaks, and retrieval is carried out for the sub-band with the highest SNR.

4. RESULTS

Fig. 3a,b,c shows how the system performs for different degrees of rotation and scaling, when n frames ($n \leq 25$) are averaged in order to improve the robustness of the system.

A threshold value of 0.025 can be observed in each figure. This guarantees a false detection probability better than 10^{-8} when the correlation peak exceeds the threshold (Fig. 3d). The value was experimentally derived for a set of 3 test sequences and a wide range of scaling and rotation attacks: the pdf of the peaks was computed for each case and the worst-case scenario determined. The resulting pdf is not Gaussian, but by fitting a zero-mean Gaussian distribution with the same standard deviation as the experimentally determined pdf, is possible to determine the optimum threshold for a given false error probability. The Gaussian distribution fits very well the worst-case scenario pdf in the zone of interest (at the extremities), and is actually chosen to be quite pessimistic. The results presented in Fig. 3d suggest that the worst-case scenario is when the sequence is marked with the correct mark.

For a false detection probability of 10^{-8}, the proposed system is invariant to scaling in the range –50% to 180%, invariant to rotation up to 70^0, and invariant to any arbitrary aspect ratio changes in the range –100% to 200% on both axes. When rotation is combined with scaling, up to 120% scaling and up to 20^0 rotation can be tolerated.

Furthermore, the system is virtually invariant to any shifting, cropping, or combined shifting and cropping. Even under severe cropping (when the useful image is only 200x200) the capacity is approximately 1500 bits/frame with turbo coding, reducing to 850 bits/frame without coding.

The system can cope very well with MPEG2 compression. The results under MPEG2 compression are presented in Fig. 4. Combined attacks like MPEG2 compression plus arbitrary frame shifts can be handled as long as the MPEG2 compression is at least 3-4Mbps. The original frame size is 720x576 (ITU-R 601).

5. CONCLUSIONS

Robustness to geometric attack is one of the most important requirements for a watermarking system. To combat this type of attack, an approach based on the FMT and log-polar/log-log representations of the video frames has been developed. This is combined with the advantages of DWT, HVS-based marking, and turbo coding to produce a very robust, high capacity video watermarking system. It outperforms many current schemes in terms of geometric invariance and channel capacity. For a wide range of attacks, the system presented in this paper meets and even exceeds the EBU watermarking recommendations [1].

6. REFERENCES

[1] L. Cheveau, E. Goray and R. Salmon, "EBU Technical Review – March 2001"

[2] B.S. Reddy & B.N. Chatterji, "An FFT-Based Technique for Translation, Rotation & Scale-Invariant Image Registration", *IEEE Trans. Image Processing*, Vol.5, No.8, August 1996.

[3] J.J.K. O'Ruanaidh, T. Pun, "Rotation, scale and translation invariant spread spectrum digital image watermarking", *Signal Processing* 66 (3), pp. 303-317, 1998.

[4] C-Y. Lin, M. Wu, J.A. Bloom, I.J. Cox, M.L. Miller, I.M. Lui, "Rotation, Scale and Translation Resilient Public Watermarking for Images", *Security & Watermarking of Multimedia Contents II, Proc. SPIE*, Vol.3971 (2000), pp.90-98.

[5] A.B. Watson, G.Y. Yang, J.A. Solomon, and J.D. Villasenor, "Visibility of Wavelet Quantization Noise", *IEEE Trans. Image Proc.*, Vol.6, 1997.

[6] J.D. Villasenor, B. Belzer and J. Liao, "Wavelet Filter Evaluation for Image Compression", *IEEE Trans. Image Proc.*, Vol.4, No.8, August 1995.

[7] C.V. Serdean, M. Tomlinson, J.G. Wade, M.A. Ambroze, "Protecting Intellectual Rights: Digital WM in the Wavelet Domain", *IEEE Int. Workshop "Trends & Recent Achievements in IT"*, Cluj-Napoca, Romania, 16-18 May 2002. (Invited Paper)

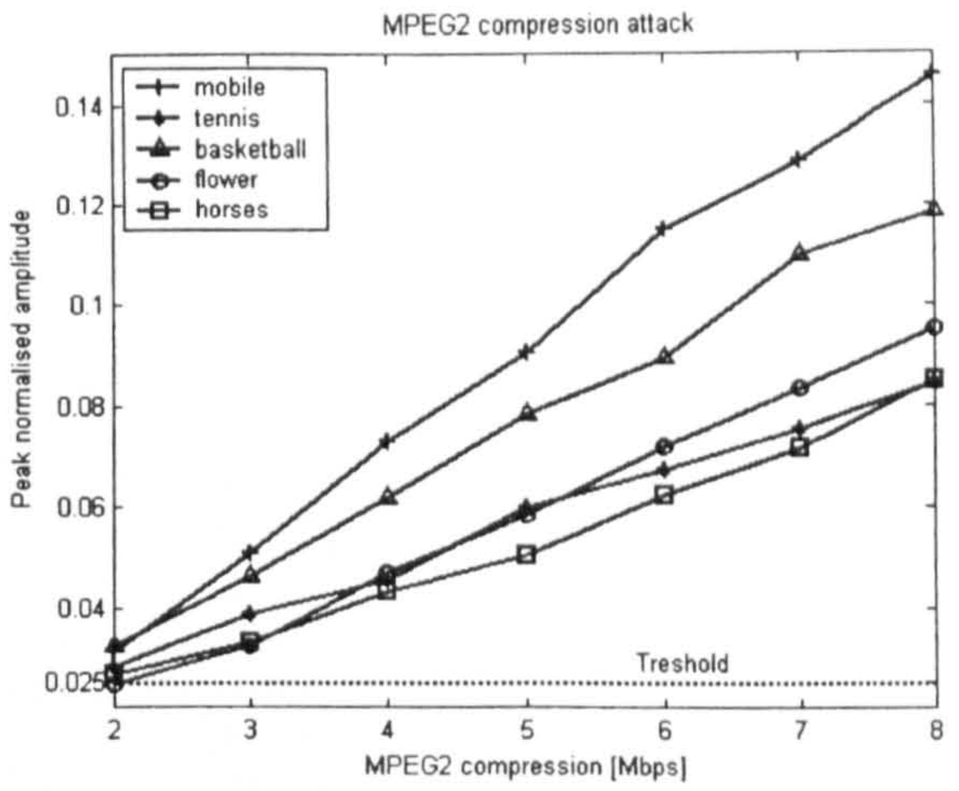
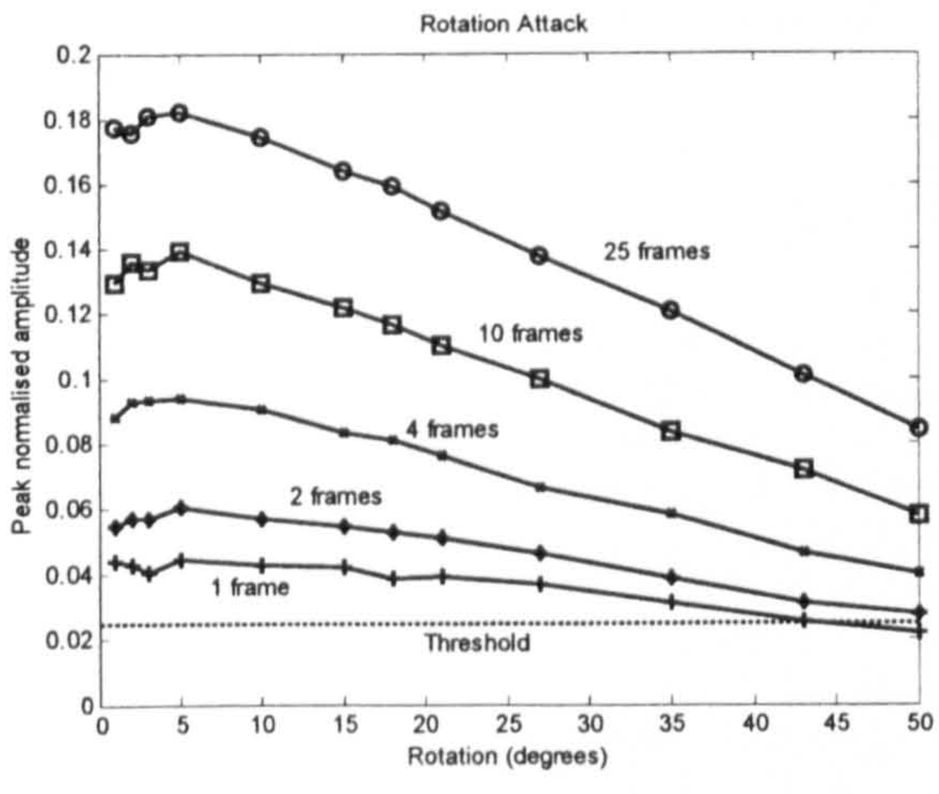
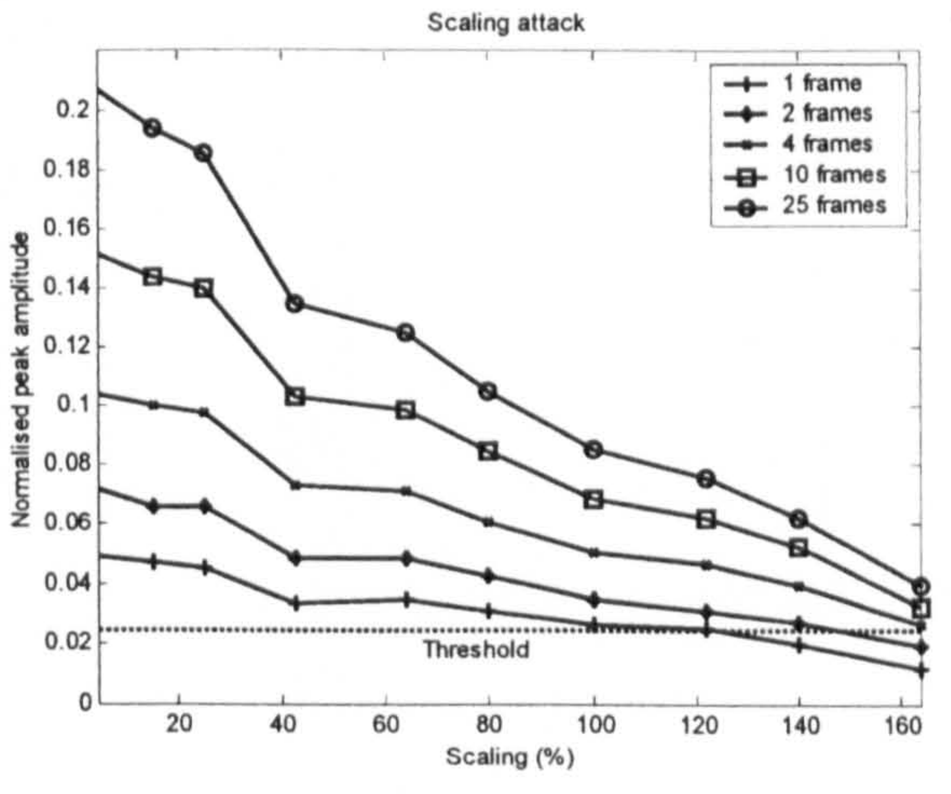


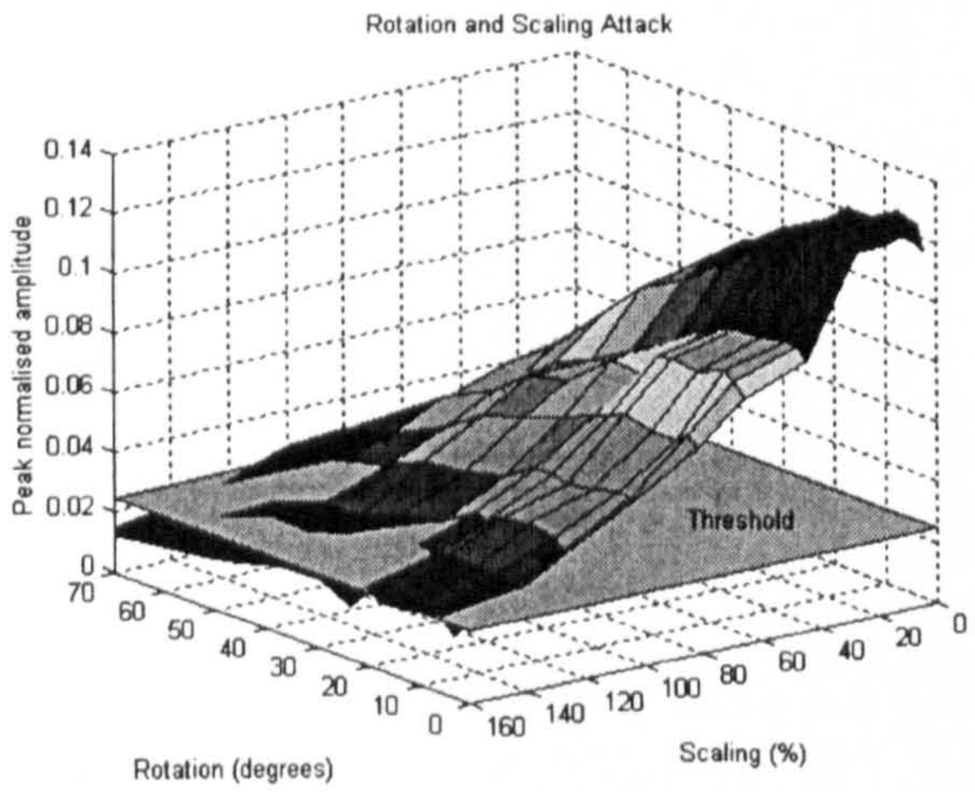
Figure 4. Performance of the system for MPEG2 attack



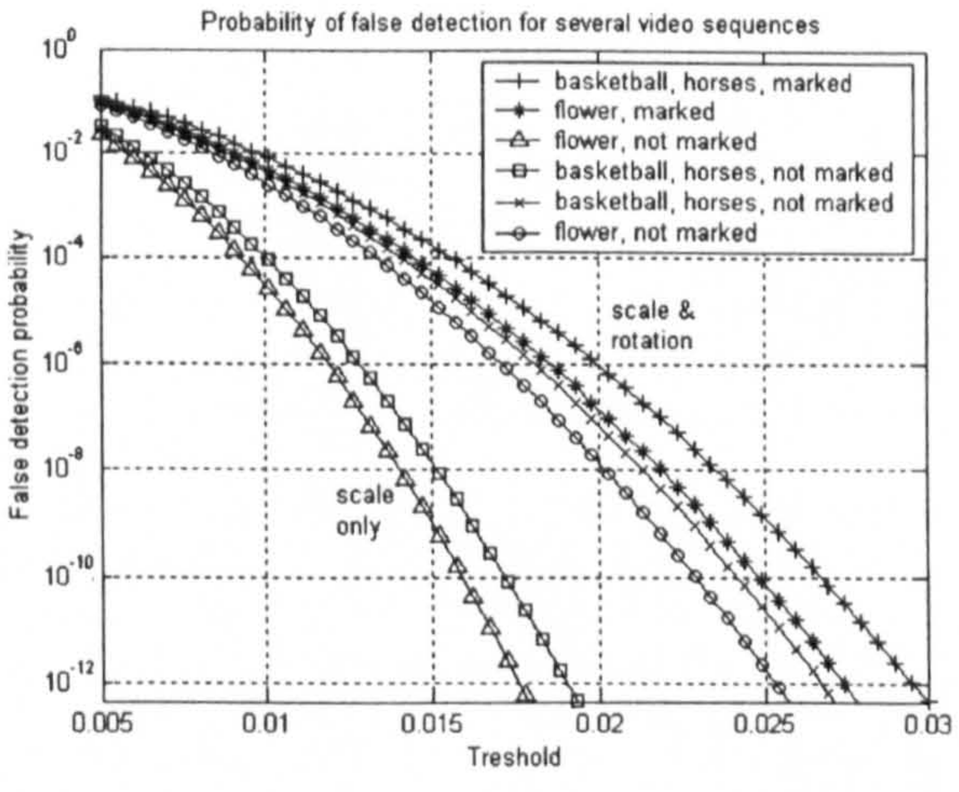
(a)



(b)



(c)



(d)

Figure 3. Performance of the system for (a) rotation and (b) scaling for basketball sequence when averaging frames; (c) Rotation combined with scaling attack (25 frames avg.); (d) Threshold selection for a desired probability of false detection.

PROTECTING INTELLECTUAL RIGHTS: DIGITAL WATERMARKING IN THE WAVELET DOMAIN

Cristian V. SERDEAN[1], Martin TOMLINSON[1], Graham J. WADE[2] and Adrian M. AMBROZE[1]
[1] University of Plymouth, United Kingdom and [2] The University of Newcastle, Australia

Abstract: The objective of the paper is to analyse the advantages and disadvantages of spatial, DFT, DCT and DWT domains and highlight the advantages offered by watermarking wavelet coefficients rather than the DCT or FFT coefficients. The reasons for the DWT advantage are analysed and the choice of a particular wavelet basis is explained. As an illustration of these advantages, the paper presents a high capacity blind video watermarking system, which embeds the data payload in the wavelet domain. In this paper the video sequence is regarded as a noisy communications channel, and the multi-bit watermark as the hidden message. In order to maximize the information capacity in the presence of attacks, the payload is embedded according to a HVS model, and is protected by state-of-the-art error correction (Turbo codes). It is shown that the DWT is significantly more robust to scaling and cropping, and gives a useful capacity improvement under a compression attack.

Keywords: copyright protection, video watermarking, spread spectrum, wavelet transform

I. INTRODUCTION

Nowadays virtually all multimedia production and distribution is digital. The advantages of digital media, for creation, processing and distribution are all well known: superior quality, more quicker and easier to edit and modify, possibility of software processing rather than the more expensive hardware alternative (if the real time processing is not a requirement), and maybe the most important advantage is the unlimited copying of digital data without any loss of quality whatsoever. This latter advantage is not desired at all by the media producers and content providers, in fact is perceived like a major threat, because it may cause them considerable financial loss.

Once the digital technology is widely available to the public, the piracy suddenly becomes a major issue. This generates the need for protecting the copyrighted material against piracy. Some typical examples are the recent court battles between the music industry and Napster, Kazaa and Morpheus. The movie and music industry are particularly keen to develop any system which will stop users copying the digital media especially now, after the introduction of Internet sharing technologies which allow users from the entire planet to share any kind of digital media between them (like Napster, Gnutella, Morpheus and many others).

In an attempt to stop this trend, the recording industry recently introduced a copyright protection system for the audio CD's which actually tries to prevent the users from copying their own legitimate CD's, and even playing these CD's on a computer. This protection system deliberately introduces during the fabrication process a substantial number of errors on the disk, in fact so many, that even the powerful error correction capability of the computer drives is defeated. This is a rather "sad" method which destroys the very core of the digital technology, lowering not only the quality, but also the reliability of the disk.

Unlike this "crude" method, digital watermarking is an unobtrusive way of protecting such material and for audio, images and video it operates by hiding a perceptually invisible signal into the host signal.

II. WATERMARKING METHODS FOR UNCOMPRESSED VIDEO

To main methods are currently used for embedding a watermark into digital media. The first method, less used is the quantisation watermarking. The second method – by far the most popular one, due to its major advantages – is the spread spectrum watermarking.

Spread spectrum radio techniques have been developed for military applications, since mid 1940's for their anti-jamming and low-probability-of-intercept properties. They allow the reception of radio signals that are over 100 times weaker than the atmospheric noise.

Moreover, the spread spectrum techniques are offering a good flexibility and are very suitable for watermarking due to the similarities between the watermarking and spread spectrum communications. The digital watermarking can be seen as a hidden communication system, in which the original image plays the role of the channel noise and attackers may try to disrupt the transfer of information. In both cases the channel is a very difficult one characterised by high levels of noise. The large bandwidth required by a spread spectrum technique is not a problem, since usually

the video sequences are quite big, offering a large number of coefficients and therefore the chip rate is sufficiently high for obtaining a robust watermarking system. The noise like spread spectrum signal is very difficult to detect/intercept and jam and is obviously spread in the entire video sequence, therefore suggesting a good robustness to certain attacks and a very secure system. Furthermore, the system can be relatively easy implemented, the watermark embedding and retrieving are based on secret keys and the system doesn't require the presence of the original video for watermark retrieving. The secret key is used for generating the same PN sequence for both embedding and retrieving. The spreading is achieved by multiplying this PN sequence with the data payload. As a result each watermark data bit is randomly spread in the entire video sequence, with a chip rate c_r. Typical for a video watermarking system, the recovery of the mark is blind, e.g. without resorting to the original video. The watermark is recovered by using cross-correlation methods, in the form of a correlation receiver of a matched filter, following the principle of optimum reception.

The uncompressed video, as found in TV studios is described by the ITU-R 601 standard. The video sequences are in raw $Y\text{-}C_B\text{-}C_R$ format. Only the luminance component Y is marked. The chrominance components are not robust at all, because they can be easily discarded, without affecting the video quality in any other way except the resulting black and white picture. Anyway marking the chrominance components has several other disadvantages. The human eye is much more sensitive to slight colour changes compared to slight luminance changes. As a result, these components have to be more lightly marked (with reduced amplitude) and from this reason are less robust compared with the luminance. Moreover, the complexity of the algorithm which uses the chrominance components is more than double, while the gain is quite small and it could be even zero if an attacker decides to discard the chrominance components. This is a strong enough reason to avoid the marking of chrominance components. Maybe in the applications where the real time requirement is not important and the cost can be tolerated one could use them in order to get a bit more robustness.

III. SPATIAL DOMAIN WATERMARKING TECHNIQUES

The first attempts to watermark an image/video sequence were done in the spatial domain. The main advantage of watermarking in the spatial domain is simplicity. Therefore the implementation time is shorter, hardware requirements are much reduced and in terms of execution time, usually the algorithms are quicker than those designed in frequency domain. Obviously this has DSP implementation advantages, being much easier to design a real-time system. Because of the lack of good visual models for spatial domain, one has to use rather empirical models as a replacement.

In terms of watermark capacity, the spatial domain is the worst place to insert a high capacity watermark. Usually, the frequency domain offers higher capacity and better robustness to attacks.

IV. WATERMARKING IN THE DFT DOMAIN

From all important frequency domain methods, the Fourier transform is the less used one. Probably the most important advantage of the DFT is its shift (translation) invariance. In other words, cyclic shifts of the video frame in spatial domain do not affect the magnitude of the DFT coefficients and therefore a watermark embedded in the magnitude of the DFT coefficients will be shift invariant. This is a highly desirable property since eliminates the need of a computationally expensive 2-D sliding window correlator.

On the other hand, due to its complex nature, the DFT offers the possibility of watermarking either the magnitude or the phase of the DFT coefficients. The phase is far more important than the magnitude of the DFT values for the intelligibility of an image, so embedding a watermark in the most important component of an image is very good since any attempts of removing the watermark will lead to heavy artefacts. Moreover, as known from the communication theory, the phase modulation often possesses superior noise immunity in comparison with amplitude modulation.

Unfortunately, in practice watermarking the phase of the DFT coefficients gives only modest results, and is very susceptible to attacks. Experiments show that the phase is quite sensitive to JPEG and MPEG attacks.

One major disadvantage of both phase and magnitude marking is the fact that in order to obtain a real image after the IDFT, one has to preserve complex conjugate symmetry of the DFT coefficients.

Changes in magnitude must preserve the positive symmetry of the Fourier coefficients:

$$F(k1,k2) = F^*(N1-k1, N2-k2) \tag{1}$$

and changes in phase must preserve the negative symmetry of the Fourier coefficients:

$$\begin{aligned} \measuredangle F(k1,k2) &\leftarrow \measuredangle F(k1,k2) + \delta \\ \measuredangle F(N1-k1,N2-k2) &\leftarrow \measuredangle F(N1-k1,N2-k2) - \delta \end{aligned} \tag{2}$$

These symmetry requirements are basically halving the watermarking space and therefore the capacity, being a serious drawback.

Another disadvantage of the Fourier domain is the lack of HVS (Human Visual System) models.

Although watermarking in Fourier domain is relatively seldom, the FFT transform is present in many watermarking systems in one way or another. For example due to its shift invariance, the FFT transform is often used to implement fast cross-correlators. According to the convolution theorem, correlation in spatial domain is equivalent with convolution in the FFT domain, and vice versa. Since a sliding correlator (e.g. a cross-correlator which is able to search for the right position of the watermark in an attacked image) is very

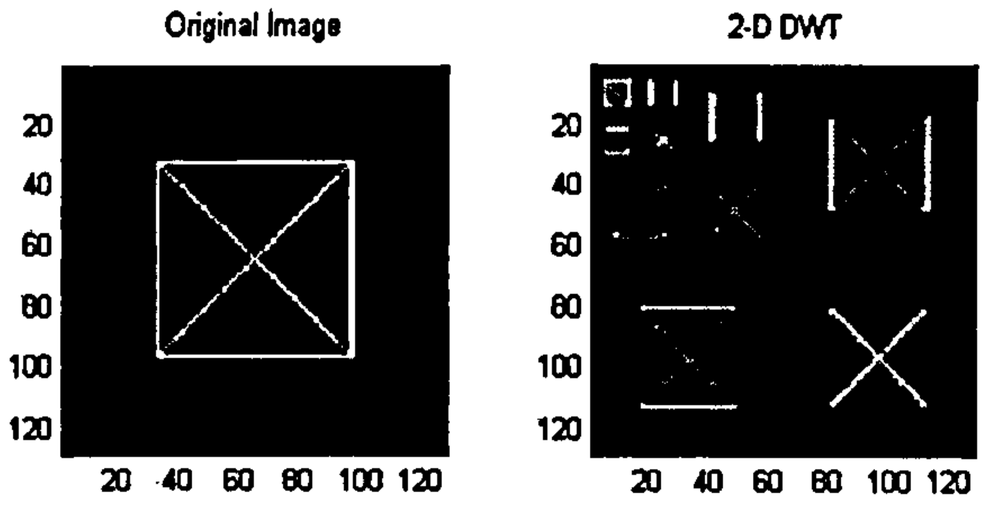


Figure 1 The 2-dimensional DWT: the original and $\lambda = 3$ levels of decomposition

computationally expensive, the efficiency of the FFT correlators is particularly welcomed. An example of such a correlator is the SPOMF (Symmetrical Phase Only Matched Filter) which is often used in image processing and pattern recognition.

Another particular area which involves the use of Fourier transform is the case of RST (Rotation, Scaling and Translation) invariant watermarking schemes.

V. WATERMARKING IN THE DCT DOMAIN

The DCT domain is far the most popular one, from several reasons. One reason is that all the major compression techniques were developed in the DCT domain (JPEG, MJPEG, MPEG1, MPEG2, H26x) and therefore the image processing community was familiar with it. Much research was carried out in developing various perceptual models for the DCT domain, and these models could be easily applied to watermarking, since watermarking and compression are very closely related. Since the compression algorithms are well known, one could compensate for it during the watermark embedding process, making the algorithm robust against compression. Furthermore marking in the frequency domain rather than spatial domain has few advantages: better robustness against certain attacks, higher capacity, more close to the HVS and relatively good frequency localisation of the coefficients. Those who are marking in the bit-stream domain (MPEG2) have the additional advantage of the direct bit-stream marking, without decoding and re-encoding the signal.

VI. WATERMARKING IN THE WAVELET DOMAIN

At the moment, the most advanced choice from all the frequency domain methods is the DWT. The advantages of the wavelet transform are presented in the following sections.

A. Multiresolution Property

The DWT is a hierarchical transform (unlike the FFT and the DCT) and offers the possibility of analysing a signal at λ different resolutions or levels (λ integer). Such multiresolution analysis gives a frequency domain representation as a function of time (or space in the 2-D case) i.e. both time/space and frequency localisation. In order to achieve this, the analysing functions must be localised in time. Formally we refer to scale and resolution, where, for the dyadic case, scale is defined as $a = 2^{\lambda}$ and resolution as $r = \frac{1}{a} = 2^{-\lambda}$. The greater the resolution, the smaller and finer are the details that can be analysed. For the 1-D case, a certain wavelet is defined by the mother wavelet function $\Psi(x)$ and a scaling function (or father wavelet) $\Phi(x)$, where the analysing wavelets are scaled and translated versions of the mother wavelet:

$$\frac{1}{\sqrt{a}}\Psi\left(\frac{x-b}{a}\right) \tag{3}$$

Defining translation $b = ka$, (k, λ integer) the dyadic case becomes:

$$\Psi_{\lambda,k}(x) = 2^{-\frac{\lambda}{2}}\Psi(2^{-\lambda}x - k)$$
$$\Phi_{\lambda,k}(x) = 2^{-\frac{\lambda}{2}}\Phi(2^{-\lambda}x - k) \tag{4}$$

For a signal $f(x)$ a wavelet coefficient is then defined as:

$$C(\lambda,k) = \int_{-\infty}^{\infty} f(x)\Psi_{\lambda,k}(x)\ dx \tag{5}$$

For the 2-D case, we have one scaling function $\Phi(x,y)$ and three wavelet functions $\Psi_{\theta}(x,y)$, where θ denotes orientation.

Different orientations extract different features of the frame, such as vertical, horizontal, and diagonal information, Fig.1. Generally speaking, edges and textures will be represented by large coefficients in the high frequency sub-bands, and are well localised within the sub-band. The use of the DWT for spread-spectrum based image/video watermarking is indicated in Fig.3 for $\lambda = 3$, and is discussed later.

B. Wavelet selection

In practice wavelet analysis is performed using multilevel filter banks. Essentially this comprises a succession of filtering and sub-sampling operations and has been widely described in the literature [2, 3, 4, 6].

For watermarking we need to select an appropriate wavelet or basis. Most of the basis development has taken place in the context of image compression [4], and fortunately watermarking and compression have many things in common. On the other hand, we certainly need to choose a basis that offers compact support. The smaller the support of the wavelet, the less nonzero wavelet coefficients will correspond to an edge for example, so basically the transform compacts more energy in the high frequency sub-bands [5]. Also we are restricted to a class of either orthogonal or bi-orthogonal wavelets. To narrow the choice even more, filter regularity, symmetry and a smooth wavelet function are important for the reconstructed image

quality. In addition, we need a reasonably good HVS model for the selected basis. Finally, for watermarking we ideally would like shift invariance in order to handle geometric attacks.

For this work we selected the Antonini 7.9 wavelet (Fig. 2), this being one of the best wavelets available for image compression [2, 3, 4]. Its important properties are highlighted below:

- Bi-orthogonal wavelet, with compact support, symmetric
- Good regularity (each filter has 2 factors [1+Z]) and the lpf and hpf are quite similar
- Simple filters (only 7 and 9 taps) with linear (zero) phase
- Shift invariant at level 1 (from the energy point of view)
- HVS model available [5]
- Smooth wavelet function

This wavelet is widely used in image compression algorithms (EZW, SPIHT), and is used in the FBI fingerprint compression standard.

C. Advantages for Watermarking

The basis function for the DFT ($f(x) = \exp(i\omega x)$) or DCT (infinite cosine) has perfect localisation in frequency but is not time/space localised. In contrast, wavelets offer a trade-off between time/space and frequency/scale, and so a watermarking scheme based on the DWT will produce a watermark with both spatially local and spatially global support (see Fig.1). This localisation makes a wavelet based scheme more robust than the DCT scheme, given geometric attacks such as cropping and scaling.

For instance, in the case of cropping, the lower frequency levels will be affected more than the high frequency ones, because of the fact that the watermark from the higher levels corresponds to a smaller spatial support. Looked at in the frequency domain, cropping corresponds to convolving the frequency components with a *sinc* function, where the width of the main lobe is inversely proportional to the width of the cropped window size [7]. This will affect all the frequency components of any scheme based on a global transform, but since the wavelet scheme has a watermark with local spatial support, the watermark will be unaffected by the cropping.

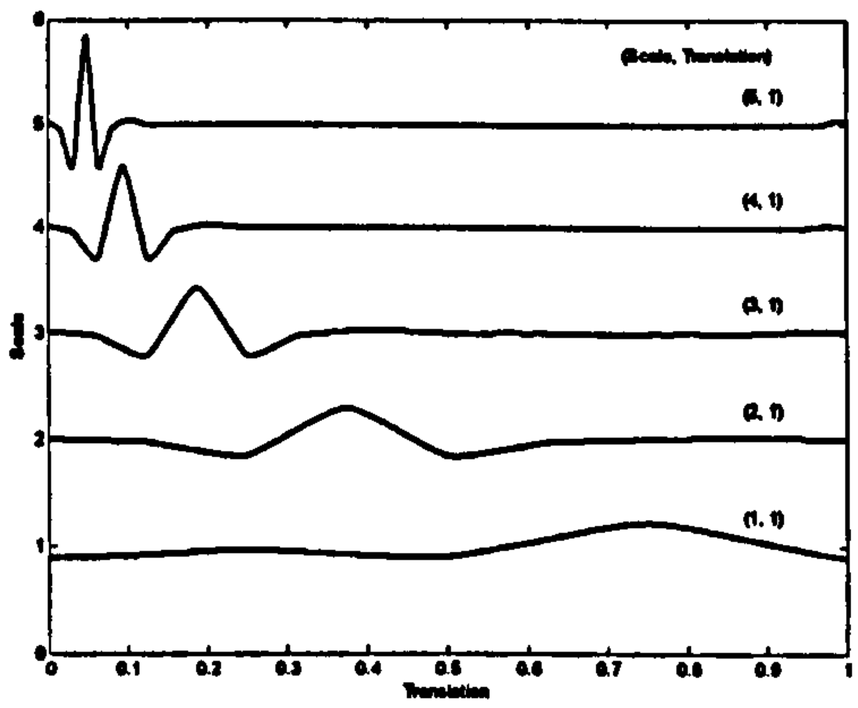


Figure 2 The Antonini 7.9 wavelets at various scales (same translation).

For scaling, because the DWT coefficients are localised both in space and frequency, whilst the DCT coefficients are only localised in frequency, it is likely that this kind of attack will be less serious for a DWT scheme. Simulation confirms this to be the case. Finally, the global spatial support of a DWT scheme will tend to be robust to operations such as low pass filtering/compression (which attenuate high frequency levels).

Another fundamental advantage of the DWT lies in the fact that it performs an analysis similar to that of the HVS. The HVS splits an image into several frequency bands and processes them independently. In a similar way, the DWT permits the independent processing of different sub-bands without significant perceptible interaction between them. Again, this is because the analysing functions Ψ are localised in space, being zero outside a space domain U i.e. the signal values located outside of domain U are not influencing the values of the coefficients within U. Similarly, if Ψ is translated to position b, the wavelet coefficient will analyse the signal around b. This local analysis is specific to the compact support wavelets. Basically for a small scale, a local analysis is performed whilst for a large scale we have a global analysis. Fig.2 shows how the wavelet functions change for different scales.

Finally, more general advantages of the DWT are:

- It is not a block based transform, and so the annoying blocking artefacts associated with the DCT are absent.
- Its multiresolution property offers more degrees of freedom compared with the DCT.
- Lower computational cost than the FFT or DCT: $O(n)$ instead of $O(n\log(n))$, where n is the order of the transform input vector.
- Better energy compaction than both the FFT and DCT in the sense that it is closer to the optimal Karhunen-Love transform.

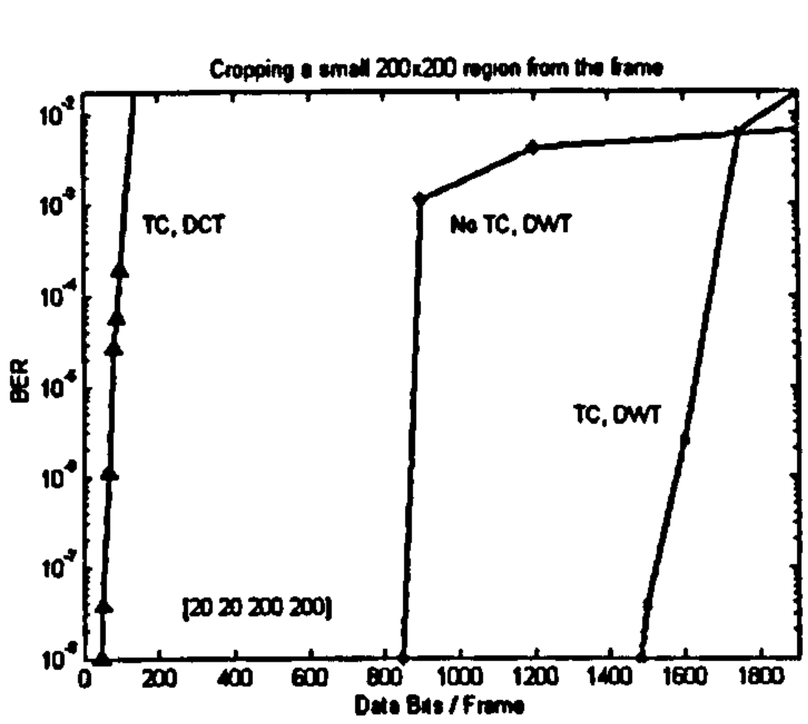
VII. THE WAVELET-BASED WATERMARKING SCHEME

Watermark embedding and the corresponding retrieval are shown in Fig.3. We have chosen to use 3 levels of decomposition. As for DCT systems, embedding uses the spread-spectrum approach and retrieval is via cross-correlation (matched filtering). The interleaver uses a separate key to that of the PN sequence in order to enhance system security and provide a random distribution of the data bits within each sub-band. Here we are exploiting the hierarchical nature of the DWT by choosing to insert a self-contained watermark in each sub-

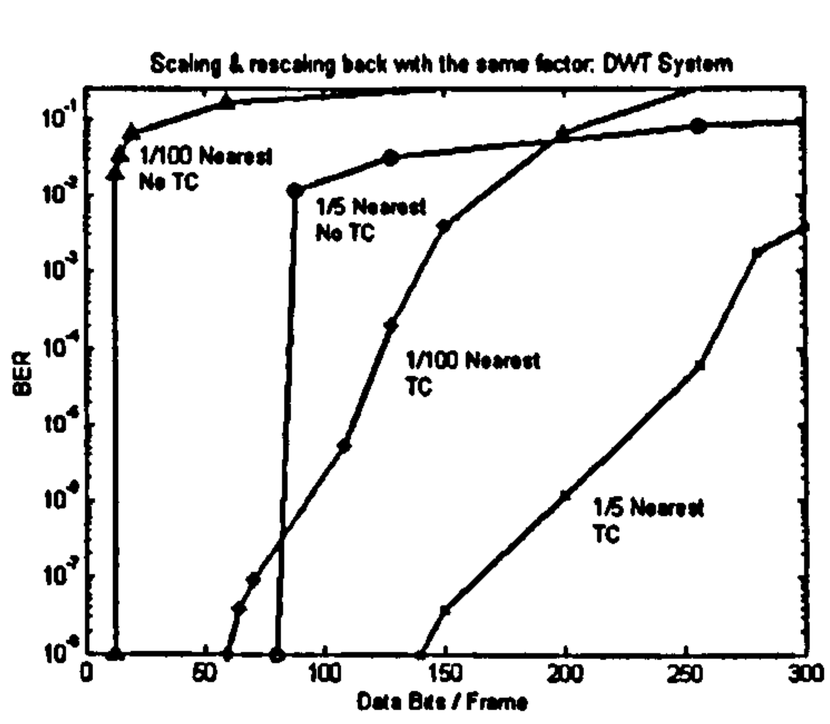
artefacts.

The DCT scheme used for comparison is the one described in [10, 11].

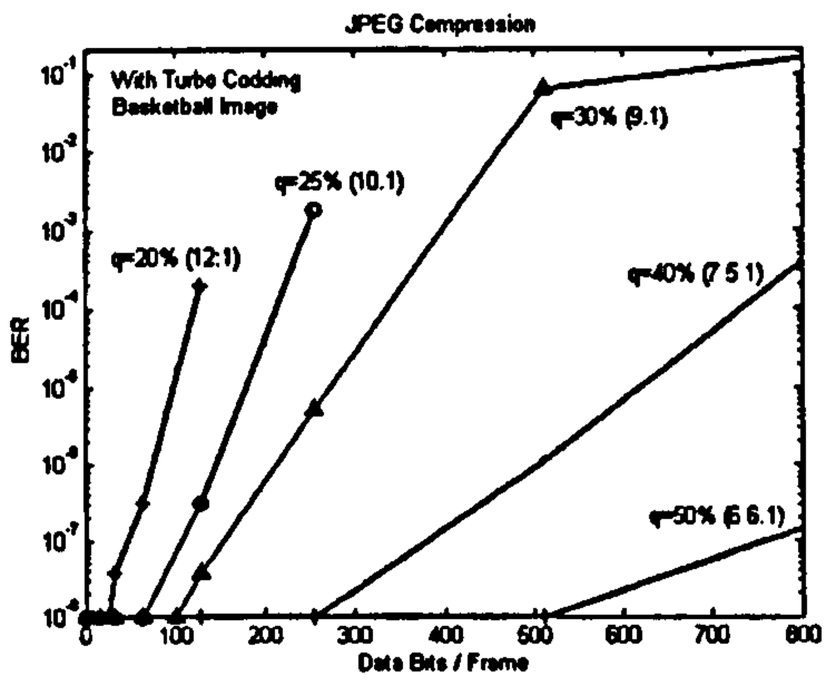
As might be expected from the compact support, the most significant advantage of wavelets occurs under cropping and scaling. For cropping, a rectangle of 200x200 pixels was selected from the upper left corner of the frame, as shown in Fig. 5(d). This location was selected since it has average detail. Clearly, cropping to this degree is an extreme case and is unlikely to occur in practice. It is apparent from Fig. 4(a) that the DCT scheme has poor performance even with FEC, whereas the DWT scheme performs very well without FEC (over 20 kbps at $BER = 10^{-3}$). With FEC the capacity increases



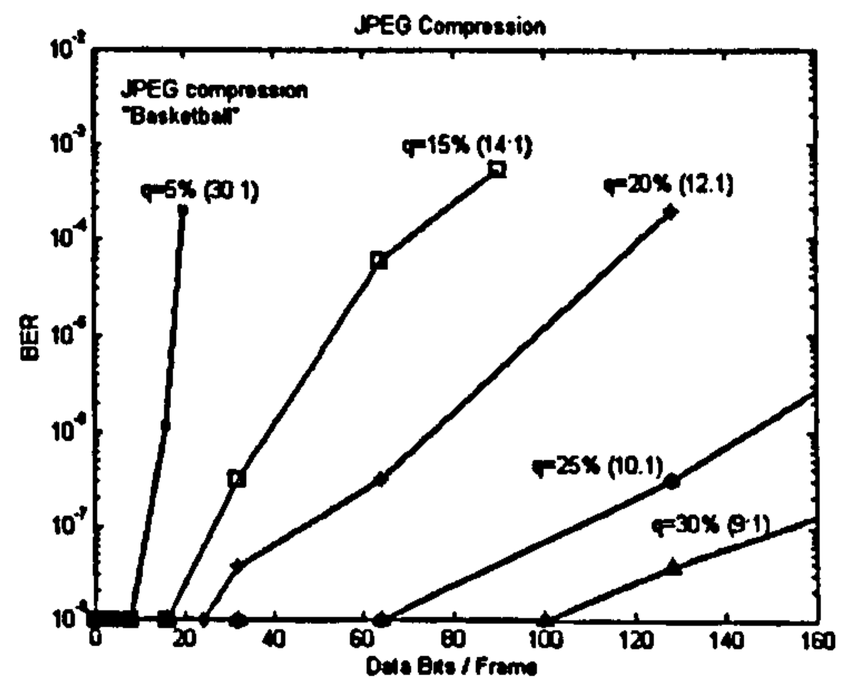
(a)



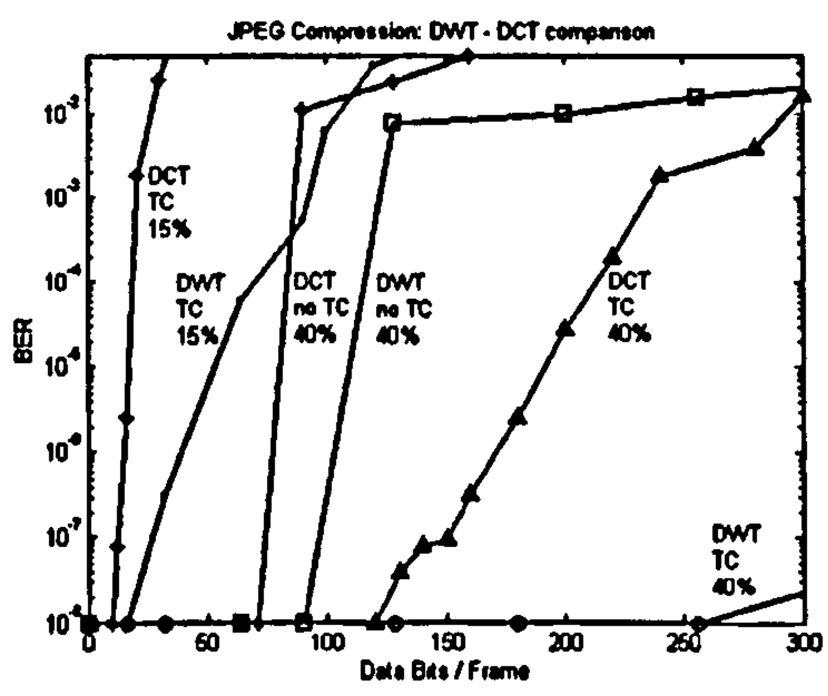
(b)



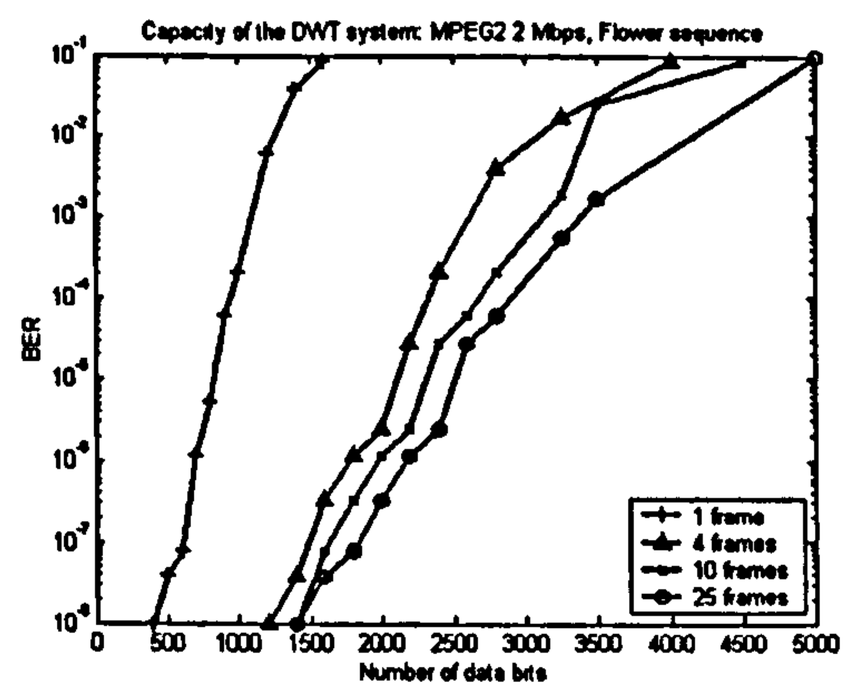
(c)



(d)



(e)



(f)

Figure 4 *Performance of the DWT system for cropping (a), scaling-rescaling (b), medium quality JPEG (c), low quality JPEG (d), DWT/DCT comparison for medium quality JPEG (e) and MPEG2 (f) attacks.*

to 37 kbps, but will reduce markedly under a combined attack.

Fig. 4(b) shows the results for scaling. The frame is scaled up or down and then brought back to the original size (720x576). Even so, with the worst kind of scaling, the DWT system performs quite well. The effect of this kind of attack results in luminosity changes and geometric distortion, Fig. 5(c). A DCT system can't cope with this attack. In contrast, the DWT gives very acceptable performance, especially when using FEC. For example, for 1/5 "nearest" scaling, the capacity is about 80 bpf (2 kbps), increasing to about 140 bpf (3.5 kbps) with FEC.

The results for JPEG compression with several different quality factors are presented in Fig. 4(c) and Fig. 4(d). As Fig. 4(d) indicates, For a relatively high compression factor of 10:1 (25% quality, slight visual artefacts) and with Turbo coding, the wavelet scheme can achieve a capacity of 64bpf (bits per frame). Even under extreme JPEG compression (30:1 compression, 5% quality, with heavy blocking artefacts) the wavelet scheme has a capacity of 8bpf. This attack is illustrated in Fig. 5(b).

A comparison of the DCT and DWT schemes under JPEG compression attack is shown in Fig. 4(e). For a quality factor of 40%, the DWT more than doubles the capacity when Turbo coding is used, the capacity being over 6 kbps at $BER = 10^{-8}$. This result clearly shows the advantage of FEC. The wavelet scheme is net superior to the DCT scheme, especially for higher quality factors and when using FEC. Again, with FEC, for a quality factor of 40% (7.5:1 compression) the capacity of the DWT scheme is double compared with the DCT scheme.

Since the capacity per frame is quite high, we can afford to increase the robustness (and the capacity as well) by inserting the same watermark in a number of n ($n \leq 25$) successive frames. In this way the recovery is much simplified since takes place only once, and is easier to combat frame dropping. This case is illustrated in Fig. 3(f) for MPEG2 compression attack, which gives an impressive capacity of about 1Kbps, when at least 4

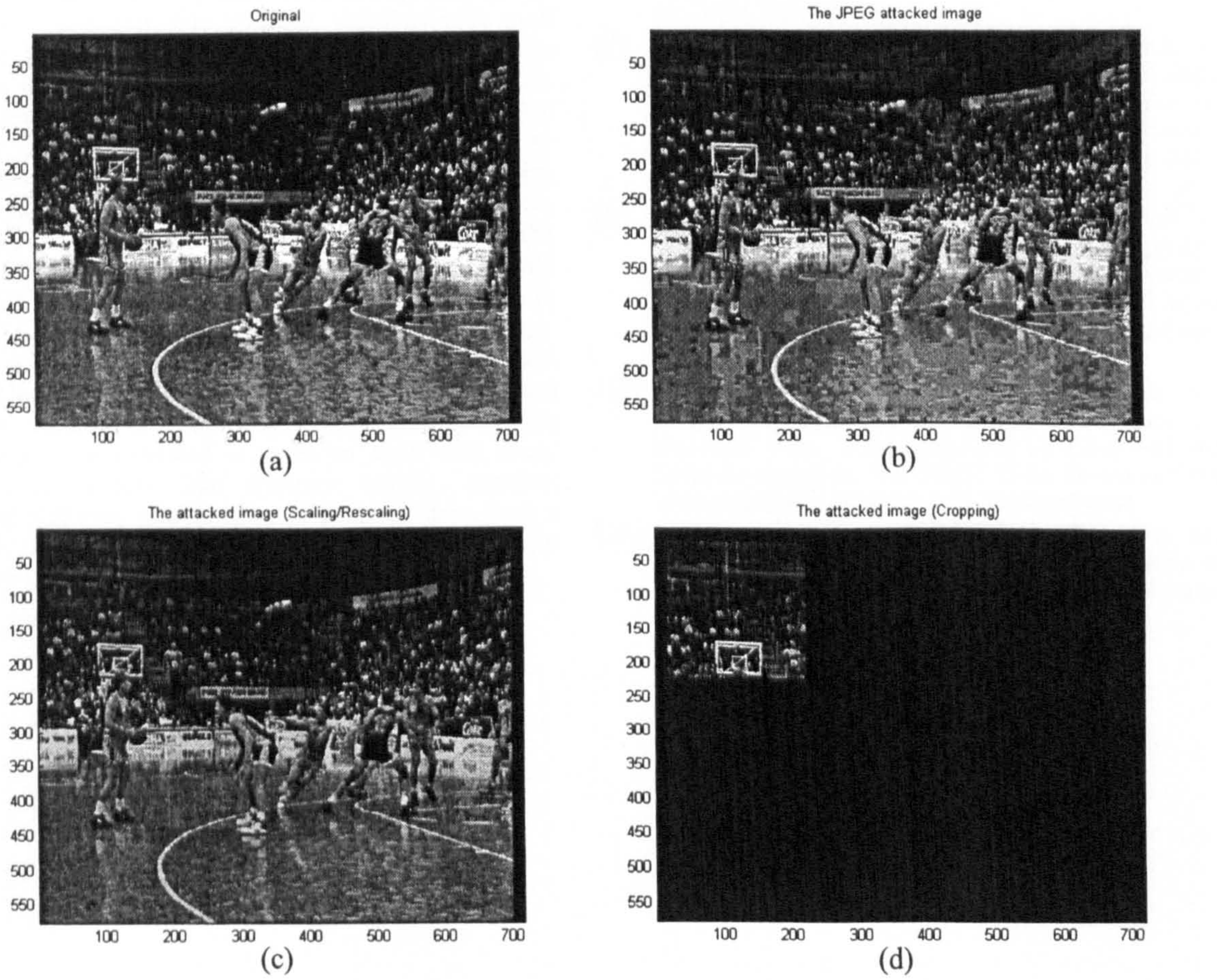


Figure 5 *A frame from the original "Basketball" sequence (a) and the effects of different attacks: (b) JPEG compression (5% quality factor, 30:1 compression ratio), (c) scaling/rescaling (1/5 and back using the 'nearest' method) and (d) cropping a small area from the original (200x200 rectangle with the upper left corner at the location [20,20])*

frames are averaged together. The improvement between the 4, 10 and respectively 25 frames averaging seems to be quite small, however this is due to the high compression applied in this case (2Mbps); for a medium level of compression the difference between these cases are much more obvious.

IX. CONCLUSIONS

The results suggest that the DWT has significant advantages under attacks which are likely to be encountered in studios e.g. compression, scaling, and cropping. Under a compression attack, the DWT can more than double the capacity of a DCT system. Under a typical scaling/re-scaling attack, a Turbo coded DWT scheme can yield capacities in excess of 1 kbps, whilst under the same conditions a DCT scheme fails. The DWT scheme has been found to be particularly robust to cropping: for example, the Turbo coded DWT scheme had a capacity of some 37 kbps, compared to 1 kbps for the DCT scheme.

The improved robustness of the DWT scheme is mainly attributed to the spatially local and spatially global support of wavelets. For example, wavelets with local support are less likely to be affected by cropping, compared to the theoretically infinitely long basis functions used in Fourier analysis. The multiresolution feature can also be exploited to optimize retrieval, by embedding all data bits in each sub-band and measuring sub-band SNR, and it gives a fundamental advantage by performing an analysis similar to that of the HVS. The DWT also has a computational advantage compared with the DCT, it does not suffer from the blocking artefacts of the DCT, and a relatively simple HVS model may suffice.

As it is shown in [8, 9], by using a second watermark embedded in the spatial domain (who acts as a reference) and by employing image registration techniques, the system can be extended in order to cope with many geometrical attacks like arbitrary scaling, rotation, shifting, and even combinations of some of them. Such a system can successfully withstand very powerful geometric attacks [8, 9].

X. REFERENCES

[1] Cheveau, L., Goray, E. and Salmon R. 'Watermarking – Summary results of EBU tests', *EBU Technical Review* – March 2001.

[2] Kingsbury N.G. and Magarey J.F.A., 'Wavelet Transforms in Image Processing', *Proc. 1'st European Conference on Signal Analysis and Prediction*, Prague, June 24-27, 1997, pp 23-34. (Invited paper)

[3] Antonini M., Barlaud M., Mathieu P. and Daubechies I., 'Image Coding Using Wavelet Transform', *IEEE Trans. Image Processing*, Vol.1, No.2, April 1992.

[4] Villasenor J.D., Belzer B. and Liao J., 'Wavelet Filter Evaluation for Image Compression', *IEEE Transactions Image Processing*, Vol.4, No.8, August 1995, 1053-1060.

[5] Lewis, A.S. and Knowles, G., 'Image Compression Using the 2-D Wavelet Transform', *IEEE Trans. Image Processing*, Vol.1, April 1992, 244-250.

[6] Watson, A.B., Yang, G.Y., Solomon, J.A., and Villasenor, J., 'Visual Thresholds for Wavelet Quantization Error', *Human Vision and Electronic Imaging, B. Rogowitz & J. Allebach Ed., Proc. SPIE*, Vol.2657, 1996, 382-392.

[7] Podilchuck, C., & Zeng, W. 'Image-Adaptive Watermarking Using Visual Models', *IEEE Journal, SAC*, Vol.16, No.4, May 1998, 525-538.

[8] Serdean, C., Ambroze, A., Tomlinson, M. and Wade, G., 'Combating Geometrical Attacks in a DWT Based Blind Video Watermarking System', *4'th EURASIP – IEEE Int. Symposium on Video/Image Processing and Multimedia Communications - VIPromCom-2002*, 16-19 June 2002, Zadar, Croatia.

[9] Serdean, C., Ambroze, A., Tomlinson, M. and Wade, G., 'DWT Based Video Watermarking for Copyright Protection, Invariant to Geometrical Attacks', *Communication Systems, Networks and Digital Signal Processing - CSNDSP'2002*, 15-17 July 2002, Staffordshire, United Kingdom.

[10] Wade, G., Serdean, C., Ambroze, A., Borda. M., Nafornita, I., 'Watermarking Uncompressed Video: An Overview', *Proc. IEEE Symposium on Electronics and Telecommunications, 'Etc. 2000'*, 23-24 November 2000, Timisoara, Romania, Vol.1, 2-15. (Invited Paper)

[11] Ambroze, A., Wade, G., Serdean, C., Tomlinson, M., Stander, J. and Borda, M., 'Turbo Code Protection of Video Watermark Channel', *IEE Proc. Vis. Image Signal Processing*, Vol.148, No.1, February 2001, 54-58.

DWT Based High Capacity Blind Video Watermarking, Invariant to Geometrical Attacks

C.V. Serdean, M.A. Ambroze and M. Tomlinson

Department of Electronic and Communication Engineering, University of Plymouth, Plymouth, UK, PL4 8AA

J.G. Wade

Department of Electrical & Computer Engineering, University of Newcastle, Callaghan, NSW 2308, Australia

Abstract

This paper describes a high capacity blind video watermarking system invariant to geometrical attacks such as shift, rotation, scaling and cropping. A spatial domain reference watermark is used to obtain invariance to geometric attacks by employing image registration techniques to determine and invert the attacks. A second, high capacity watermark, which carries the data payload, is embedded in the wavelet domain according to a human visual system model. This is protected by a state-of-the-art error correction code (Turbo code). For a false detection probability of 10^{-8}, the proposed system is invariant to scaling up to 180%, rotation up to 70^0, and arbitrary aspect ratio changes up to 200% on both axes. Furthermore, the system is virtually invariant to any shifting, cropping, or combined shifting and cropping. The system is also robust to MPEG2 compression even when combined with shifting and cropping.

1. Introduction

One of the most difficult problems in digital video watermarking is watermark recovery in the presence of geometric attack. Typical attacks are frame shift (translation), cropping, scaling, rotation, and change of aspect ratio, and recovery is particularly difficult when these are combined together [1]. The work presented in this paper was carried out in the context of uncompressed video where geometric attacks tend to be less severe compared to those for image watermarking [1]. On the other hand, the recovery problem is compounded for video since it must be carried out blind due to the difficulty of storing the original. In this case, for the typical spread spectrum watermarking system, blind retrieval is performed via cross-correlation between the

marked video and the secret pseudo-noise (PN) sequence used to spread the watermark at the embedding stage. Recovery is straightforward given perfect synchronisation between the attacked video and the PN sequence, but is difficult when geometric attacks destroy the synchronisation. In this case it is possible to perform some form of sliding correlation in order to re-establish synchronisation i.e. multiple cross-correlations over a specified search space. Unfortunately the search space grows very quickly, making it difficult to recover the watermark in a reasonable time. Clearly, given that retrieval in a video context must be done in near real time, the computational problem is very significant in the presence of attacks.

A partial solution is to employ fast correlation. Using a Symmetrical Phase Only Matched Filter (SPOMF), as proposed in [7, 2, 3], solves the frame shift problem, leading to a shift invariant watermarking system. Unfortunately, this technique can be applied only to frame shifts.

It is well known in image processing that transformation of Cartesian coordinates into log-polar coordinates prior to the FFT gives scale and rotation invariance [4, 5]. This is the Fourier-Mellin transform (FMT), and it was first used by O'Ruanaidh [9] to achieve invariance for image watermarking. Unfortunately, marking in the FMT domain has two major drawbacks: the need to compute the inverse log-polar transform (a lossy operation that drastically reduces system performance), and the need to maintain the FFT symmetry, which halves the watermark capacity. These drawbacks make the system impractical, although an improved technique was proposed by Lin [10].

This paper combines the advantages of an algorithm based on the FMT image registration techniques, with watermarking in the Discrete Wavelet Transform (DWT) domain. The idea is to first undo geometric attack using the FMT approach and an

additional spatial reference watermark used only for registration purposes. Once the attack parameters are determined, the geometric attacks are undone and the resulting frame is passed to the main watermark decoder. The main watermark, which carries multi-bit data, is inserted in the DWT domain according to a human visual system (HVS) model. The system can be regarded as a noisy communications channel and so is protected by turbo coding. The net result is a system that can withstand severe geometric attack, the limiting attack being defined by a threshold yielding a false detection probability of 10^{-8}, and capacity being defined by a bit error rate (BER) of 10^{-8}. It offers higher capacity and robustness compared to other watermarking systems described in the literature [1].

Section 2 introduces the basis of the image registration module and its invariance to shift, rotation, scale and arbitrary aspect ratio changes, insisting on the implementation issues arising for blind watermarking. Section 3 describes the overall system and Section 4 discusses the DWT marking, HVS embedding and turbo code protection. The system's performance for different attacks is presented in Section 5.

2. Combating Geometric Attack using Log-polar and Log-log Transformation

Ideas developed by Casasent [4, 5] were later adopted for image processing in the context of image registration [6, 7, 8]. This involves two images: the original and an attacked copy, and the objective of image registration is to determine the parameters of the geometric distortion. The attack can then be inverted to give geometric alignment of two images. When registering two images, the noise is relatively small, and so the correlator usually performs very well.

However, this approach cannot be used for blind watermarking since the original video frame is not available. To overcome this problem, we suggest a "blind

registration" technique. The unavailability of the original is circumvented by using a spatial spread spectrum reference watermark. In this case the PN sequence used to embed the reference watermark plays the role of the "original image", and the attacked watermarked video frame (watermark plus significant noise, arising from the video itself) represents the "attacked image". Therefore, for "blind registration" the signal to noise ratio (SNR) is very low relative to that for image registration. In the proposed system we embed two different watermarks. The first is a 1-bit watermark used exclusively for geometric reference, and for simplicity is embedded in the spatial domain. The second, multi-bit watermark is used for the data payload, and is embedded in the DWT domain.

The desired geometric invariance can be achieved by using the FMT to convert rotation and scale to spatial shifts, which are then easily recovered by a SPOMF. Performing a log-log transform (LLT) of the input permits recovery from arbitrary scale changes. Similarly, a log-polar transform (LPT) converts rotation and scaling to spatial shifts, and permits recovery from rotation and scaling. The theory behind the LPT/LLT and its application to image registration is well represented in literature [4, 5, 6, 7, 8, 9].

Since the FMT is not shift invariant, it is necessary to apply the Fourier magnitude of the frame (rather than the frame itself) to the input of the LPT/LLT module. The Fourier magnitude is shift invariant and so the attack parameters can be found even in the presence of shift. After undoing the attack, the shift is then recovered by performing a simple SPOMF correlation. This technique works well in the particular case of image-image registration [6], since the correlation peaks are relatively large and the phase loss can be tolerated. Unfortunately, this technique fails for "blind registration" and therefore this approach cannot be used for retrieval under combined attack.

A LPT permits recovery over a wide range of scale changes, rotation, or even combined scale-rotation attack. If a LLT is used, then it is possible to recover arbitrary aspect ratio changes (different scale factors for x and y axes). The shifts alone are easily recovered using a SPOMF module. However, shift recovery from a combined attack (e.g. shift/scaling/rotation, or shift/aspect-ratio change) requires a comprehensive search for all of the possible shifts [8], and is computationally intensive.

3. A Robust System

Fig.1 shows a schematic of the proposed system. The decision block determines if the reference watermark is present (to within a desired false detection probability), and if present it automatically determines the attack parameters. One advantage of using two watermarks is now apparent: if the reference cannot be found, it is assumed that either the video is not marked, or that the mark is destroyed, and recovery of the main watermark payload is abandoned (saving computation time). Also, in the proposed scheme, the two watermarks are embedded in different domains, and each watermark is embedded at the full strength dictated by its own HVS model.

As stated, the reference watermark is used exclusively as a reference and can be regarded as just one-bit. This is embedded in the spatial domain using spread spectrum, together with a simple visual model that inserts a stronger watermark in those regions where it is less easily observed. The same reference watermark is embedded in all the frames in order to increase the robustness (the SNR at the correlator input is increased via frame averaging) and the detection speed of the algorithm (the registration process takes place only once, not for each separate frame). This is possible because attacks must be identical for each frame in order to avoid temporal artefacts. Moreover, the

three registration modules can work in parallel in order to increase the speed of the algorithm.

Fig.2 shows implementation detail of the LPT/LLT registration module. The role of the Laplacian high pass filter (HPF) is to remove low and medium frequency video components (which represent noise) and pass only the high frequency components, which contain the spread spectrum, noise-like watermark. This significantly improves the correlator performance.

4. The DWT video watermarking scheme

The wavelet watermarking offers many advantages compared with FFT or DCT watermarking, as shown in [15]. The wavelet transform itself has been widely described in the literature [11, 12, 13, 14]. For watermarking, we selected the Antonini 7.9 wavelet, as being one of the best wavelets available [11, 12, 13]. The reasons for choosing this basis and its important properties are discussed in [15].

The proposed DWT video watermarking scheme is shown in Fig.3, where the noisy channel corresponds to the video sequence. The watermark energy is maximised by embedding the main payload according to an HVS model, and the BER is minimised through the use of turbo coding. This significantly increases the operational capacity of the system [15]. In the recovery process, since the cross-correlator performs a sequence of correlation sums, it follows from the Central Limit theorem that the cross-correlation peaks have a normal distribution [16]. This is convenient for turbo decoding since the latter generally assumes a Gaussian input. Thus, for any particular system, the mean μ and variance σ^2 of the correlation peaks defines a *SNR* of the channel: $SNR = (\mu/\sigma)^2$. The corresponding *BER* for an uncoded system is simply $BER_u = Q[\mu/\sigma] = Q\left[\sqrt{SNR_u}\right]$. For a coded system, the decoded *BER* is

$BER_c = f(SNR_c)$ where f is a known function for a particular iterative decoder. For this work we used a rate ¼ multiple parallel-concatenated convolutional code (3PCCC) rather than the basic turbo code (2PCCC) in order to improve performance [16].

Watermark embedding is shown in Fig.4, where we use 3 levels of decomposition. Embedding uses the spread-spectrum approach and retrieval is via cross-correlation (matched filtering). The security of such a system relies in the secret watermarking keys, K_1 and K_2. The interleaver (key K_2) provides a random distribution of the data bits within each sub-band. Watermark retrieval is shown in Fig.5. The video sequence is filtered using a Laplacian 3x3 filter prior to cross-correlation in order to improve the performance of the correlator. It is advantageous to have a self-contained watermark (all data bits) in each sub-band, since a SNR can be determined for each sub-band as an indicator of sub-channel quality. Different types of attack affect different levels and orientations in different ways, and so it is always possible to select an optimal sub-band via SNR. Correlation is therefore performed separately for each sub-band, orientation and level, obtaining each time a set of cross-correlation peaks (one peak for each embedded data bit). A SNR is then computed for each set of cross-correlation peaks, and retrieval is carried out for the sub-band with the highest SNR.

4.1 HVS-based embedding

The hierarchical nature of the DWT is exploited by inserting a self-contained watermark in each sub-band, i.e. all payload bits are inserted into each sub-band. The watermark is embedded using amplitude modulation as follows:

$$C_i^M = \begin{cases} C_i + \underbrace{\alpha \frac{Q(\lambda,\theta)}{Q_{\min}} \cdot \frac{|C_i|}{mean(|C_i|)}}_{S} \cdot W_i, & \text{(details)} \\ \quad if \quad S > 24, \quad then \quad S = 24. \\ C_i + \alpha \frac{Q(\lambda,\theta)}{2} \cdot \frac{|C_i|}{mean(|C_i|)} \cdot W_i, & \text{(approximation)} \end{cases} \tag{1}$$

where Q_{min} is the minimum value from the quantisation matrix $Q(\lambda,\theta)$, W_i is the watermark, C_i is the original wavelet coefficient, C_i^M is the marked coefficient, λ is the level and θ denotes orientation. Note that (1) incorporates media dependence ($|C_i|$), essential for robust watermarking. The high frequency sub-bands and the largest coefficients are marked more heavily, since modification of these coefficients is less likely to incur visible artefacts. The HVS is incorporated in $Q(\lambda,\theta)$ which is computed according to [14].

5. System Performance and False Detection Probability

The registration module provides invariance to frame shift, rotation, scaling, rotation combined with scaling, and aspect ratio change. The system can also handle a range of other attacks, such as cropping, shift + cropping, MPEG compression, compression + shift + cropping. These attacks are illustrated in Fig.6, for test sequences "basketball" and "flower". For watermarking, basketball represents a typical average sequence, while flower is known to be a very difficult sequence. The invisibility of the mark was subjectively assessed using specialised hardware.

Fig.7 shows the performance of the system for different degrees of rotation, when n frames are averaged in order to improve the robustness of the system. Since the minimum watermarking segment is 25 frames, then $n \leq 25$. Compared with the $n = 1$

case, the cross-correlation peak for $n = 25$ is about four times larger. Similar results are presented in Fig.8 for scaling. Fig.9 and 10 show the system performance for $n = 25$ and different degrees of rotation and scaling. Finally, Fig.11 presents the case of rotation combined with scaling for the "basketball" sequence ($n = 25$).

A threshold value of 0.025 can be observed in each figure (Fig.7-11, 13). This guarantees a false positive detection probability better than 10^{-8} when the correlation peak exceeds the threshold. The value was experimentally derived for a set of 3 test sequences and a wide range of scaling and rotation attacks: the pdf of the peaks was computed for each case and the worst-case scenario determined. The resulting pdf is not Gaussian due to the large number of very small peak values, but by fitting a zero-mean Gaussian distribution with the same standard deviation as the experimentally determined pdf, the resulting Gaussian distribution can be used to determine the optimum threshold for a given false error probability. The Gaussian distribution fits very well the worst-case scenario pdf in the zone of interest (at the extremities), and is actually chosen to be quite pessimistic. We have investigated several hypotheses: when the sequence was marked with the correct watermark, when the sequence was not marked and when the sequence was marked with a wrong mark, for different attacks and different strength of the attacks, and finally for 3 different video sequences. The results (Fig.12) suggest that the worst-case scenario is when the sequence is marked with the correct mark, and show that the 0.025 threshold is appropriate for a false detection probability of 10^{-8}.

5.1 Scaling, Rotation, and Cropping Attack

As can be seen in Fig.9, the system is invariant to any amount of rotation smaller than 70^0. Fig.10 shows that the system can handle any degree of scaling up to

180% (it can also handle scaling up to –50% i.e. smaller frames). The system therefore exceeds the EBU recommendation [1] for both rotation and scaling.

When rotation is combined with scaling, up to 120% scaling and up to 20^0 rotation can be tolerated, even for the "flower" sequence. All simulations assume a bilinear interpolation in the log-polar module. Our tests show that bilinear interpolation leads to a substantial performance increase (almost double) compared with a simple nearest neighbour interpolation. A combined attack of 20^0 rotation plus 100% scaling is shown in Fig.6(d).

The system copes very well with cropping attack. Even under severe cropping (as in Fig.6(g), where the useful frame area is only 208x196) the capacity is approximately 1500 bits/frame with turbo coding, reducing to 850 bits/frame without coding [15].

5.2 Compression Attack

We have investigated system performance for MPEG2 and JPEG [15] compression attacks. The registration module can cope with MPEG2 compression as low as 2-3 Mbps. The system can handle even combined attacks like MPEG2 compression (as low as 3-4Mbps) combined with frame shifts, as illustrated in Fig.13. In terms of capacity, the DWT watermark survives MPEG2 compression at 2Mbps, for a capacity higher than 1200bits/second (Fig.14).

6. Conclusions

Robustness to geometric attack is one of the most important requirements for a watermarking system. To achieve this, an approach based on the image registration techniques and LPT/LLT of the video frames has been developed.

Parameter	EBU Recommendations	Proposed System
GENERAL PARAMETERS OF THE SYSTEM		
Watermarking Minimum Segment (WMS)	1sec, 5sec	min 1sec
Data Capacity	64bits/WMS	≥1200bits/WMS @ 2Mbps MPEG2
Probability for error-free payload per WMS	$>10^{-8}$	10^{-8}
False positive probability per WMS	$<10^{-8}$	$<10^{-8}$
Format of original and watermarked signals	ITU-R 601 (ITU-T BT.656)	ITU-R 601 (ITU-T BT.656)
Watermark recovery	Blind	Blind
ROBUSTNESS TO ATTACKS		
MPEG2 compression	2-6Mbps MPEG2	2-6Mbps MPEG2
Colour-space conversion	YES	Invariant
Shift	up to 320x288	higher than 320x288
Scaling	desired: 200%, -50% best achieved: 140%, -70%	180%, -50%
Aspect-ratio conversion	16:9 ↔ 4:3	16:9 ↔ 4:3 (easy) 200%, -100%
Small rotation	up to 2^0	up to 2^0
Noticeable rotation	up to 10^0	up to 70^0
Small bend/shear	up to 2^0 (10^0)	NO
Cropping	minimum size: 320x288	even smaller than 200x200
Combined attacks	NOT SPECIFIED	YES (wide range)

Table 1 The performance of the proposed system compared with EBU's recommendations.

An additional spatial reference watermark compensates for the unavailable original video sequence and makes possible the geometrical "blind registration". This is combined with the advantages of the DWT, HVS-based marking, and turbo coding to produce a very robust, high capacity video watermarking system.

In this way, the advantages of both techniques are preserved: the speed and efficiency of image registration techniques and the robustness and high capacity of the wavelet system. The performance of the proposed system, compared with the EBU recommendations [1] is summarised in Table 1.

References

1. Cheveau, L., Goray, E. and Salmon R., 'Watermarking – Summary Results of EBU Tests', *EBU Technical Review*, No. 282, March 2001.

2. Kuglin, C.D., Hines, D.C., 'The Phase Correlation Image Alignment Method', *Proc. IEEE Intl. Conference on Cybernetics and Society*, September 1975, pp. 163-165.

3. Horner, J.L., Gianino, P.D., 'Phase Only Matched Filtering', *Applied Optics*, Vol.23, No.6, 15 March 1984.

4. Casasent D., Psaltis D., 'Scale Invariant Optical Correlation Using Mellin Transforms', *Optics Communications*, Vol.17, No.1, April 1976.

5. Casasent D., Psaltis D., 'Position, Rotation, and Scale Invariant Optical Correlation', *Applied Optics*, Vol.15, No.7, July 1976.

6. Reddy, B.S., Chatterji, B.N., 'An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration', *IEEE Trans. Image Processing*, Vol.5, No.8, August 1996.

7. Chen, Q-S, Defrise, M., Deconinck, F., 'Symmetric Phase-Only Matched Filtering of Fourier-Mellin Transforms for Image Registration and Recognition', *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.16, No.12, December 1994.

8. Wolberg, G., Zokai, S., 'Robust Image Registration Using Log-Polar Transform', *Proc. Intl. Conf. On Image Processing*, Vancouver, Canada, September 2000.

9. O'Ruanaidh, J.J.K., Pun, T., 'Rotation, scale and translation invariant spread spectrum digital image watermarking', *Signal Processing*, 1998, 66, (3), pp. 303-317.

10. Lin, C-Y, Wu, M., Bloom, J.A., Cox, I.J., Miller, M.L., Lui, I.M., 'Rotation, Scale and Translation Resilient Public Watermarking for Images', *Security and Watermarking of Multimedia Contents II, Proc. of SPIE*, 2000, Vol.3971, pp.90-98.

11. Kingsbury, N.G., Magarey, J.F.A., 'Wavelet Transforms in Image Processing', *Proc. first European Conf. on Signal Analysis and Prediction*, Prague, June 24-27, 1997, pp 23-34.

12. Antonini M., Barlaud M., Mathieu P. and Daubechies I., 'Image Coding Using Wavelet Transform', *IEEE Trans. Image Processing*, Vol.1, No.2, April 1992.

13. Villasenor J.D., Belzer B. and Liao J., 'Wavelet Filter Evaluation for Image Compression', *IEEE Trans. Image Processing*, Vol.4, No.8, August 1995, 1053-1060.

14. Watson, A.B., Yang, G.Y., Solomon, J.A., & Villasenor, J., 'Visibility of Wavelet Quantization Noise', *IEEE Trans. Image Processing*, Vol.6, 1997, 1164-1175.

15. Serdean, C.V., Tomlinson, M., Wade, J.G., & Ambroze, M.A., 'Protecting Intellectual Rights: Digital WM in the Wavelet Domain', *Proc. IEEE Intl. Workshop "Trends & Recent Achievements in Information Technology"*, Cluj-Napoca, Romania, 16-18 May 2002, pp. 70-77.

16. Ambroze, A., Wade, G., Serdean, C., Tomlinson, M., Stander, J. and Borda, M., 'Turbo Code Protection of Video Watermark Channel', *IEE Proc. Vision, Image & Signal Processing*, Vol.148, No.1, February 2001, pp. 54-58.

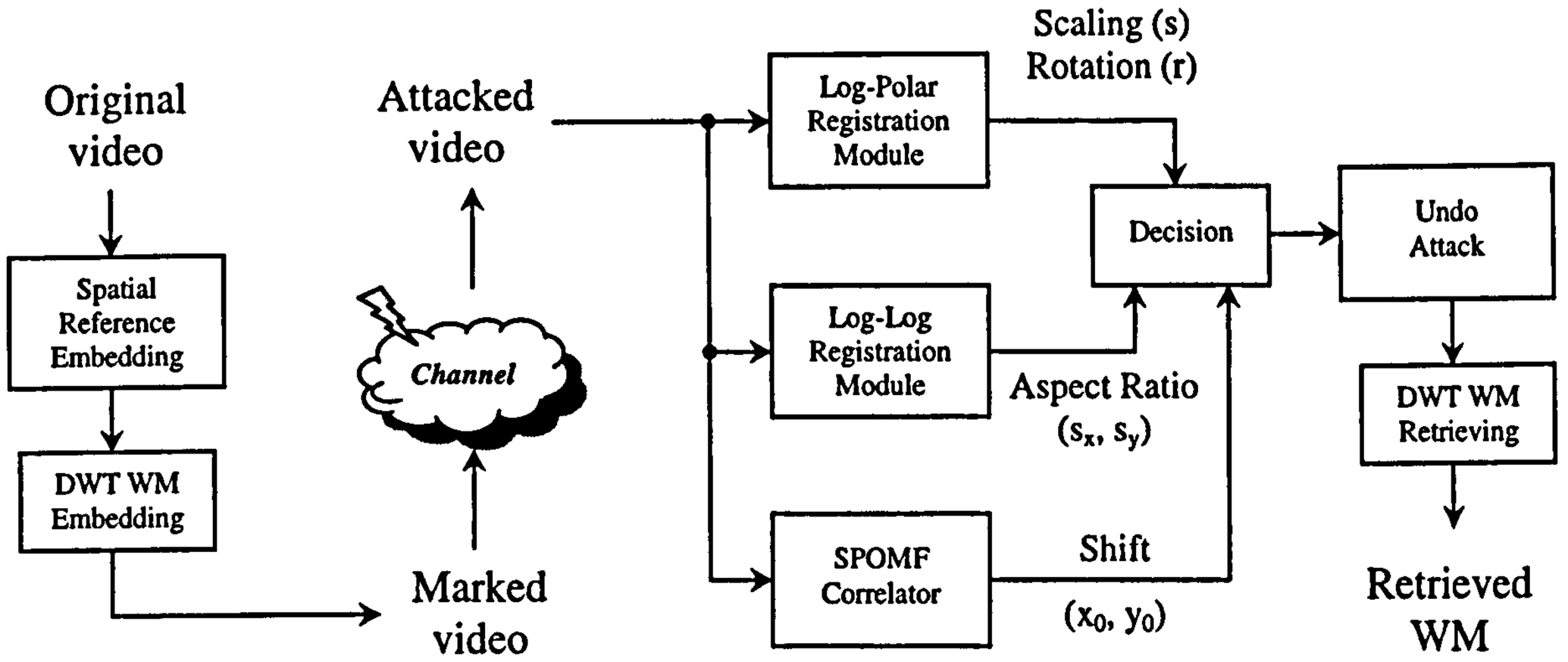


Figure 1. Block schematic of the geometric invariant video watermarking system

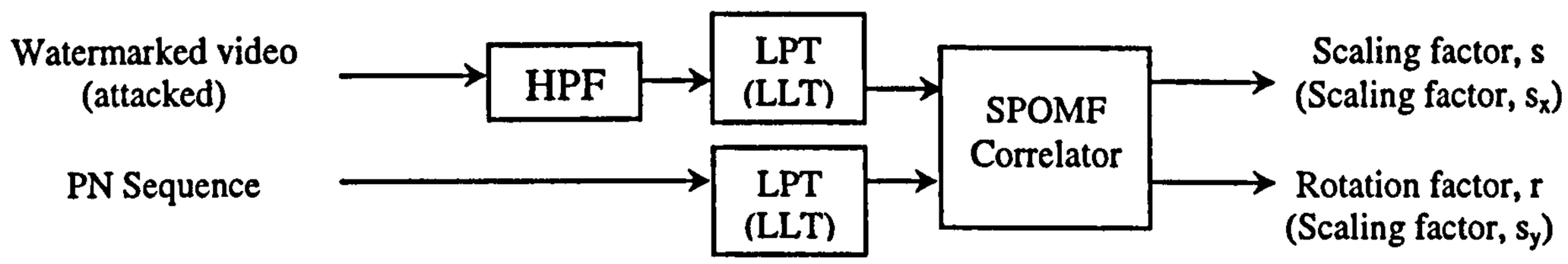


Figure 2. The log-polar / log-log module

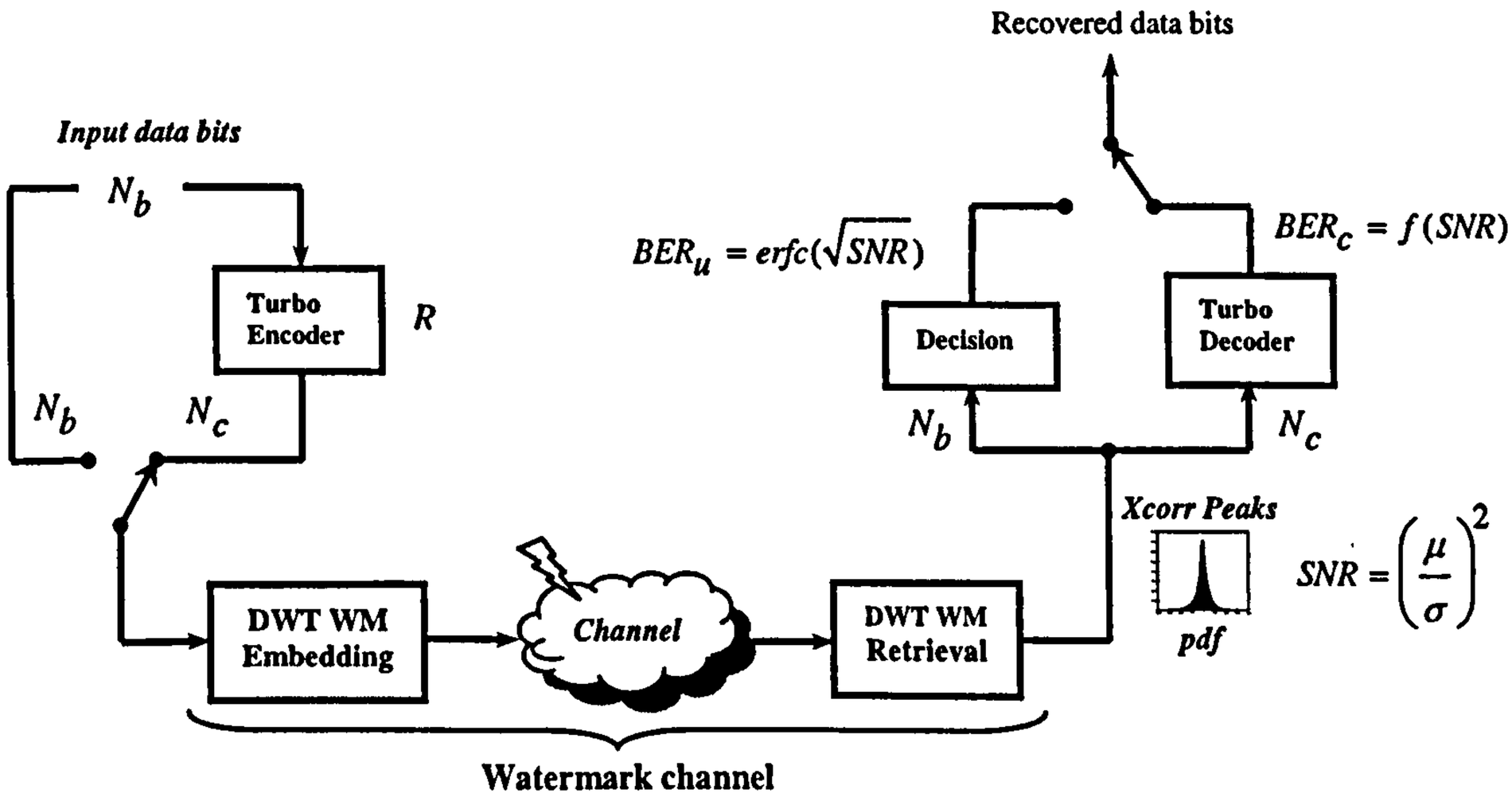


Figure 3. The DWT video watermarking scheme

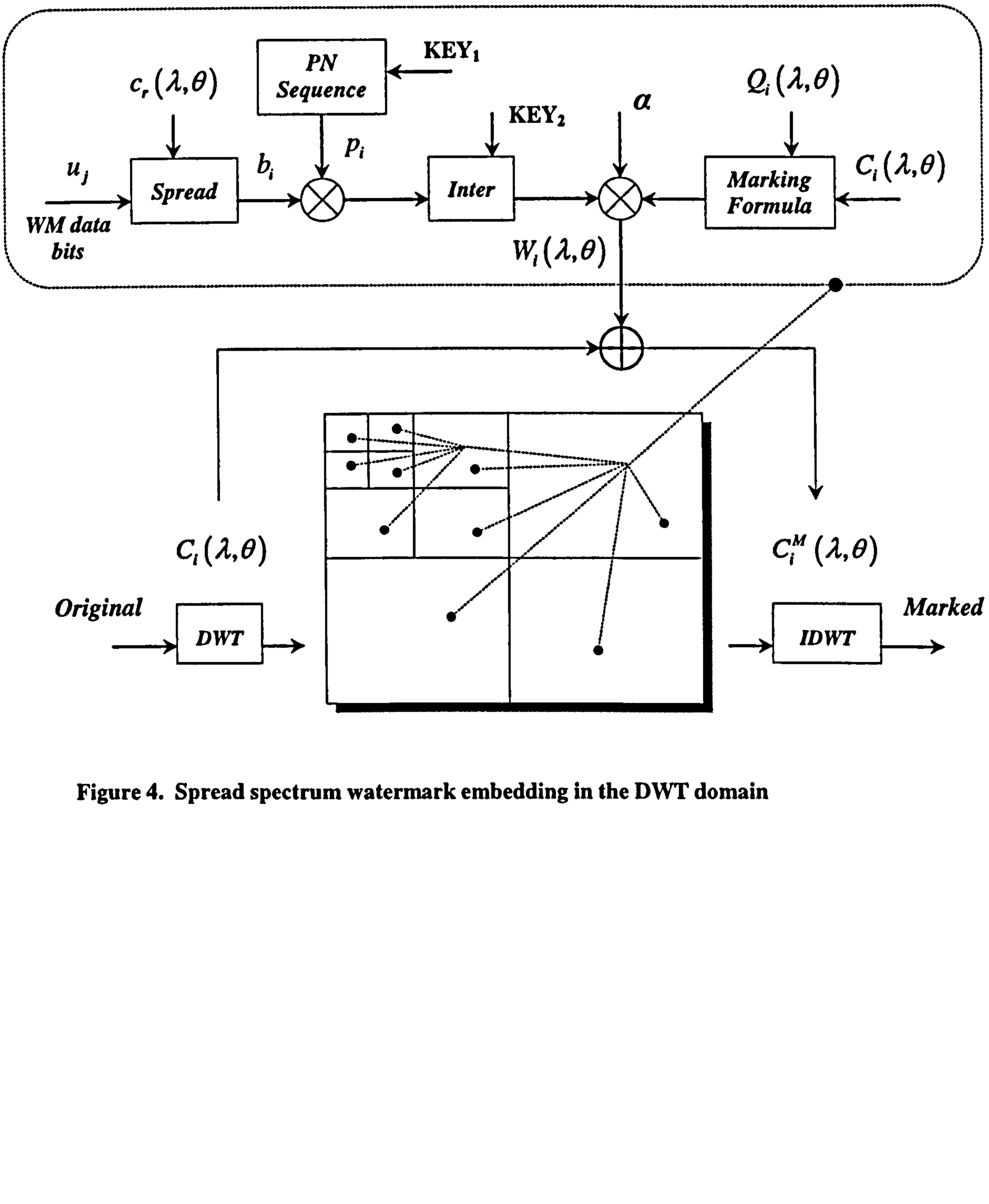


Figure 4. Spread spectrum watermark embedding in the DWT domain

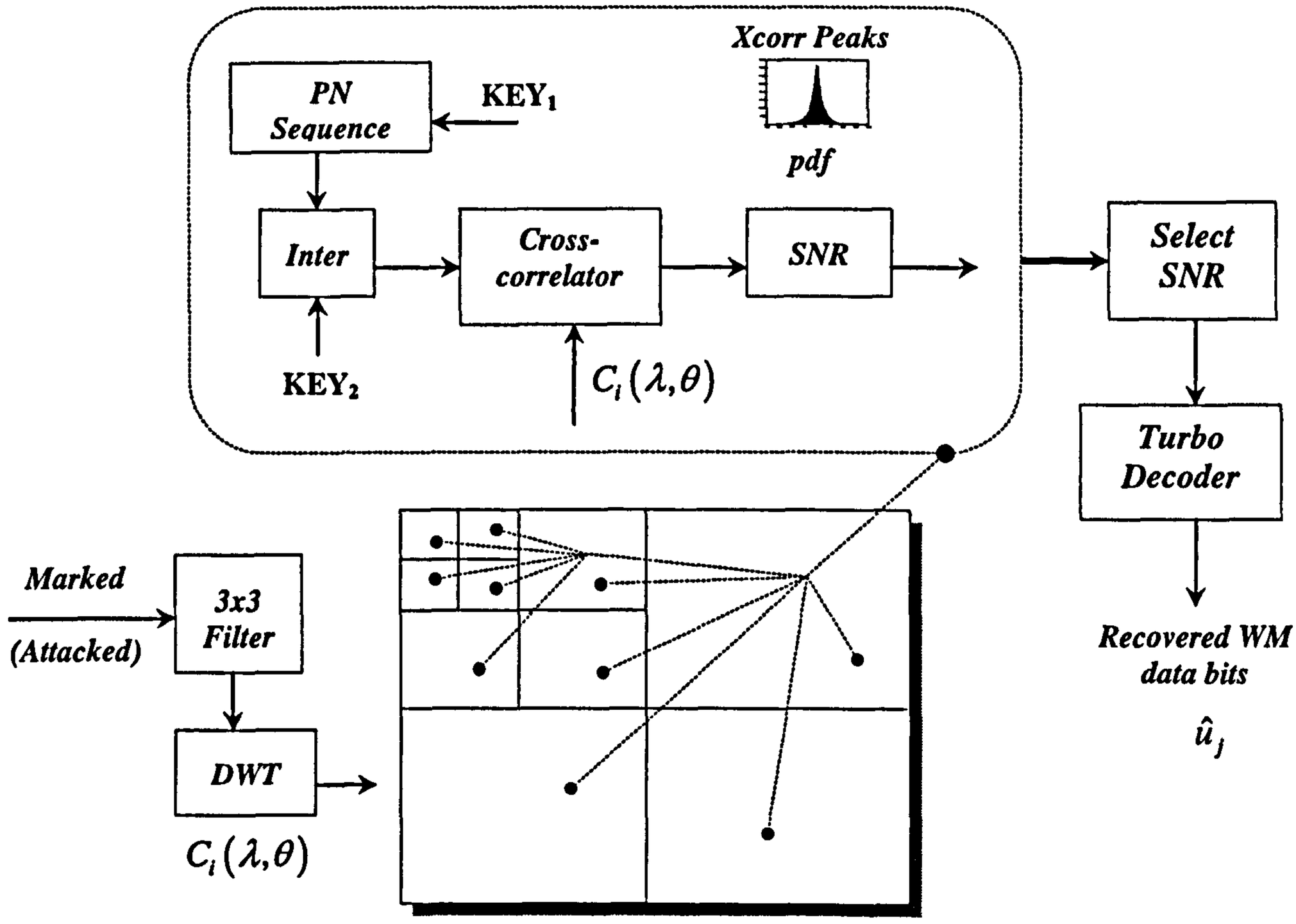


Figure 5. Spread spectrum DWT watermark retrieval

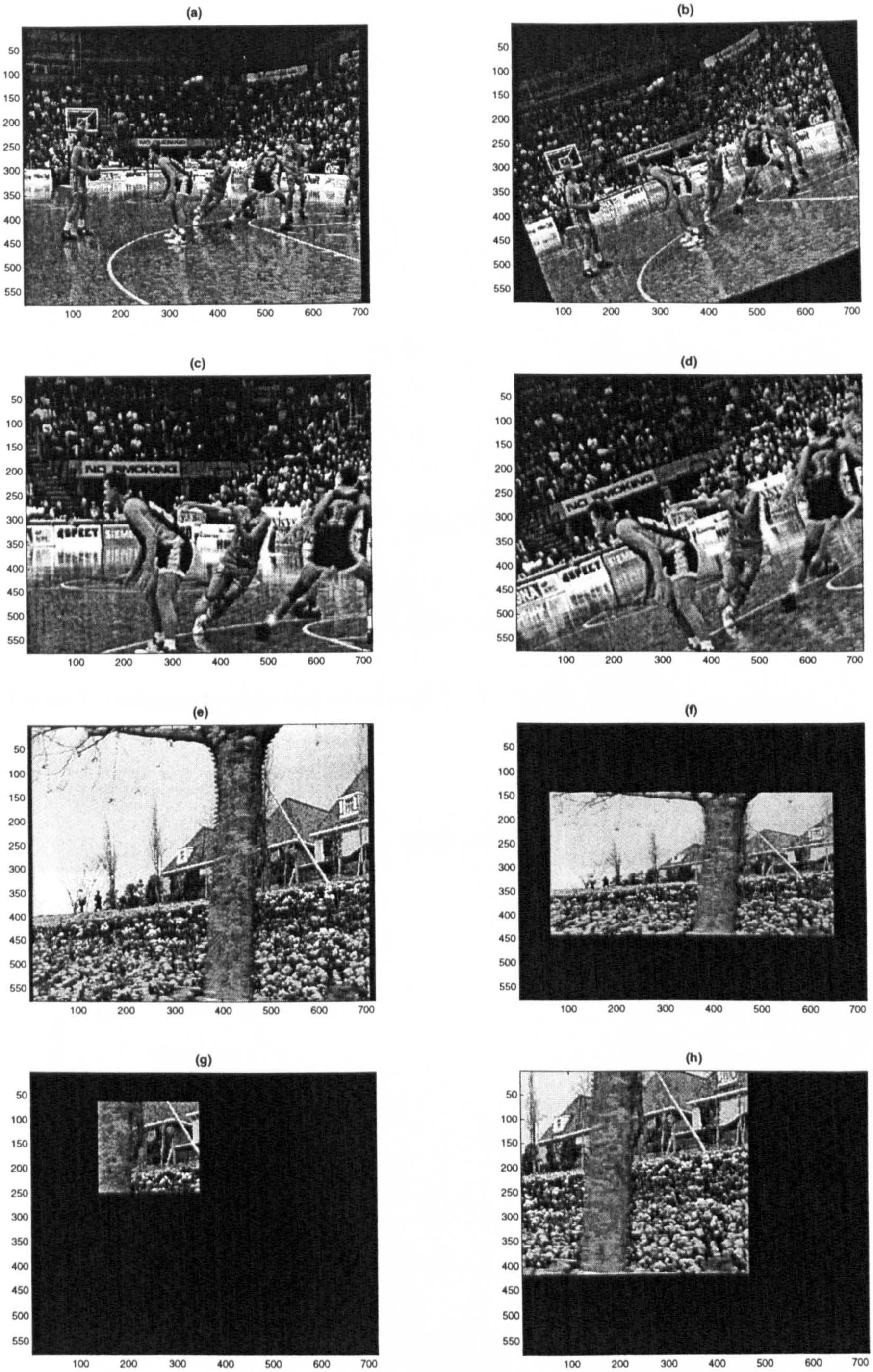


Figure 6. The effects of different attacks: (a) original basketball, (b) 20^0 rotation, (c) 100% scaling, (d) 20^0 rotation combined with 100% scaling, (e) original flower, (f) arbitrary scaling, image is rescaled from [576x720] to [300x600], (g) cropping [400,200,208,196] combined with shift [140,240], (h) MPEG2 (2Mbps) combined with shift [160,240]

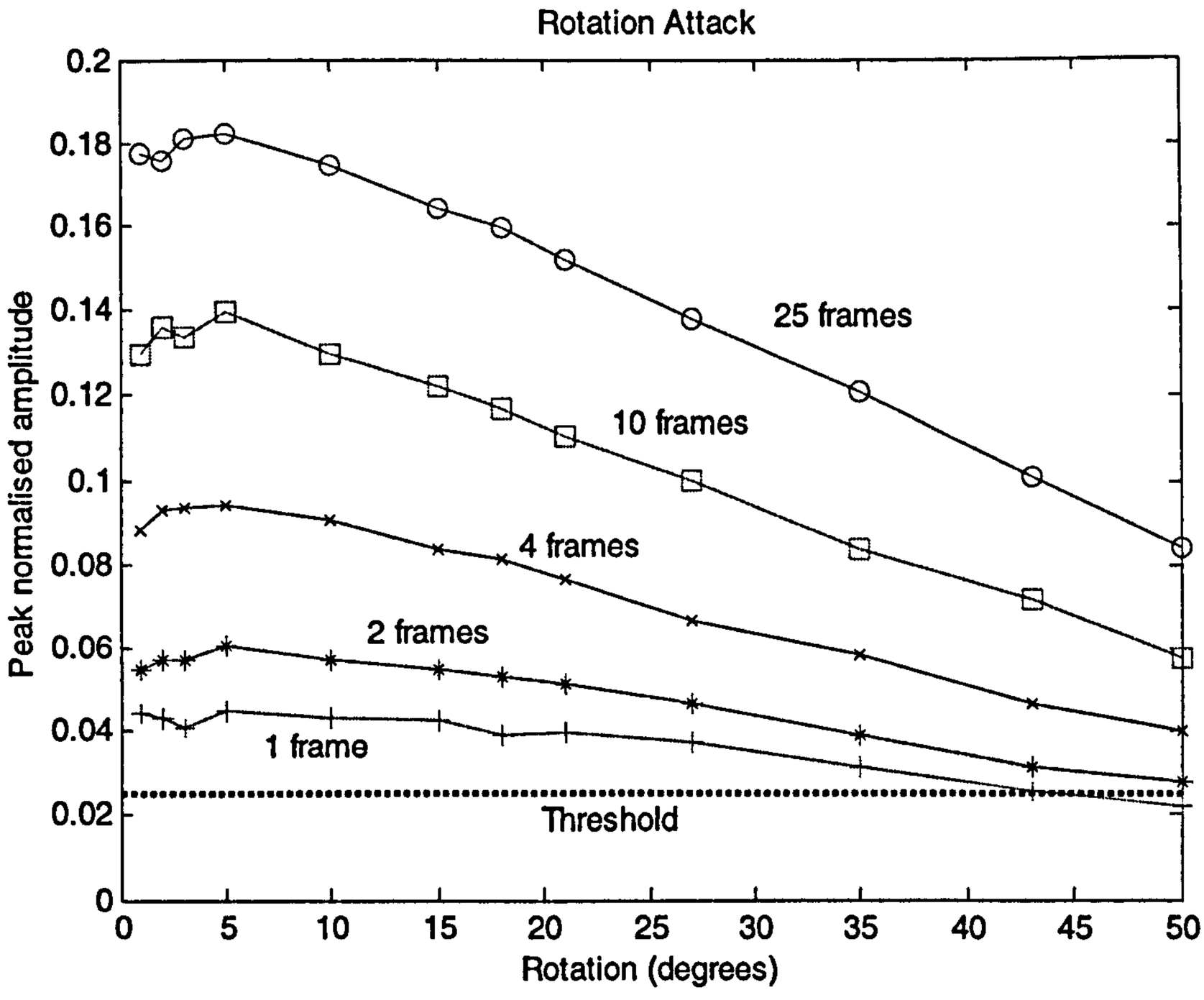


Figure 7. Performance of the system for rotation when averaging frames.

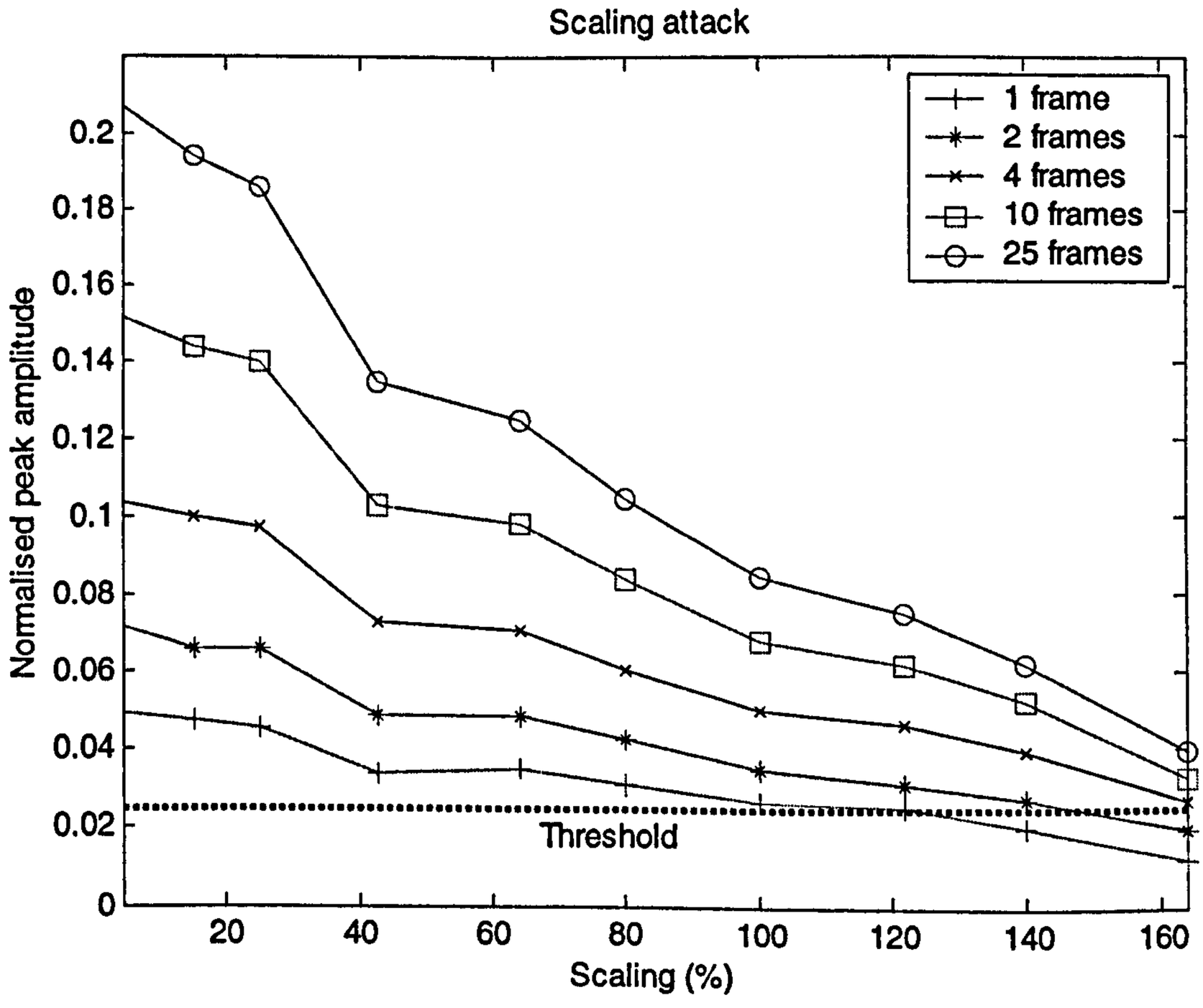


Figure 8. Performance of the system for scaling when averaging frames.

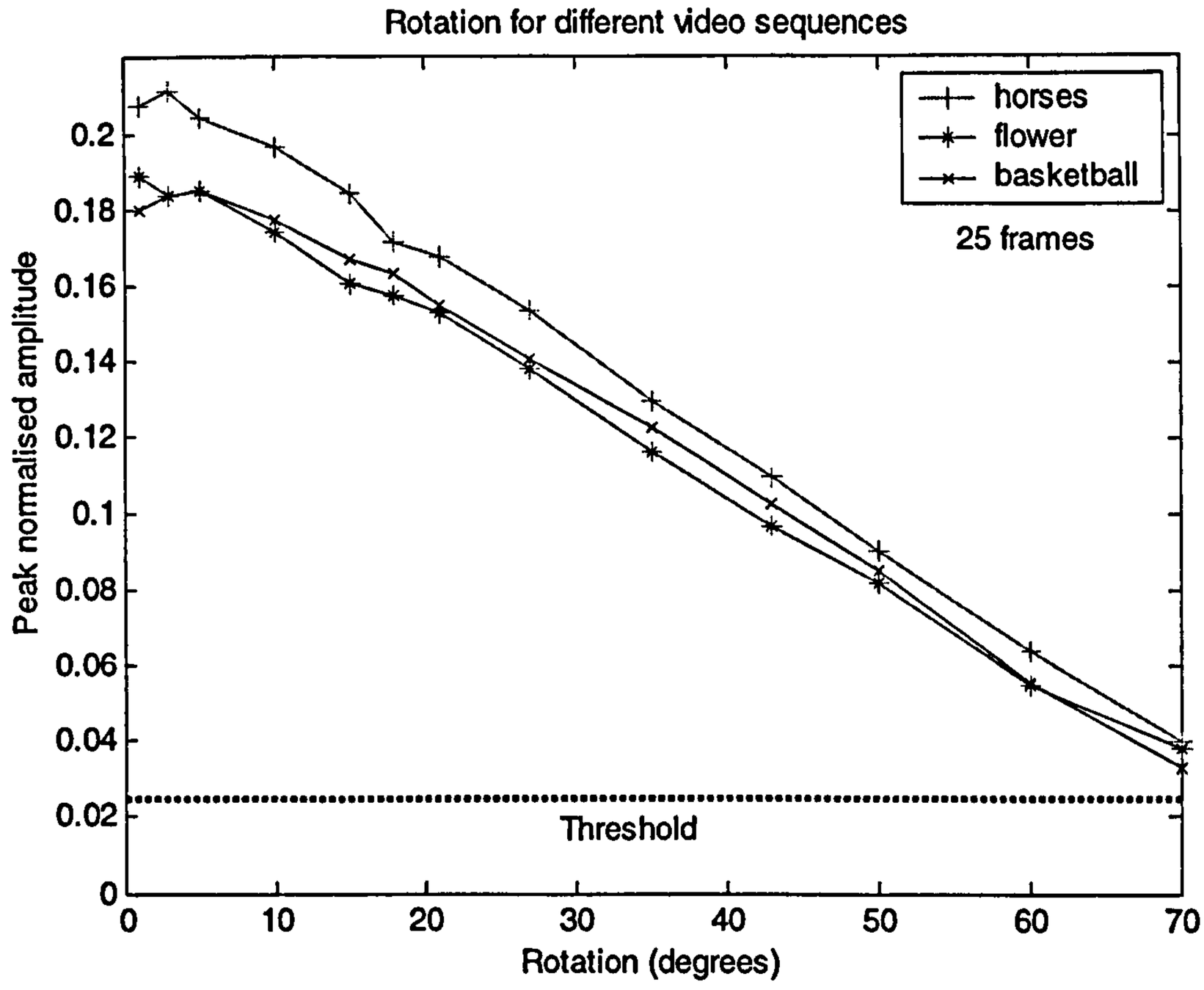


Figure 9. Peak normalised amplitude for different video sequences under rotation attack

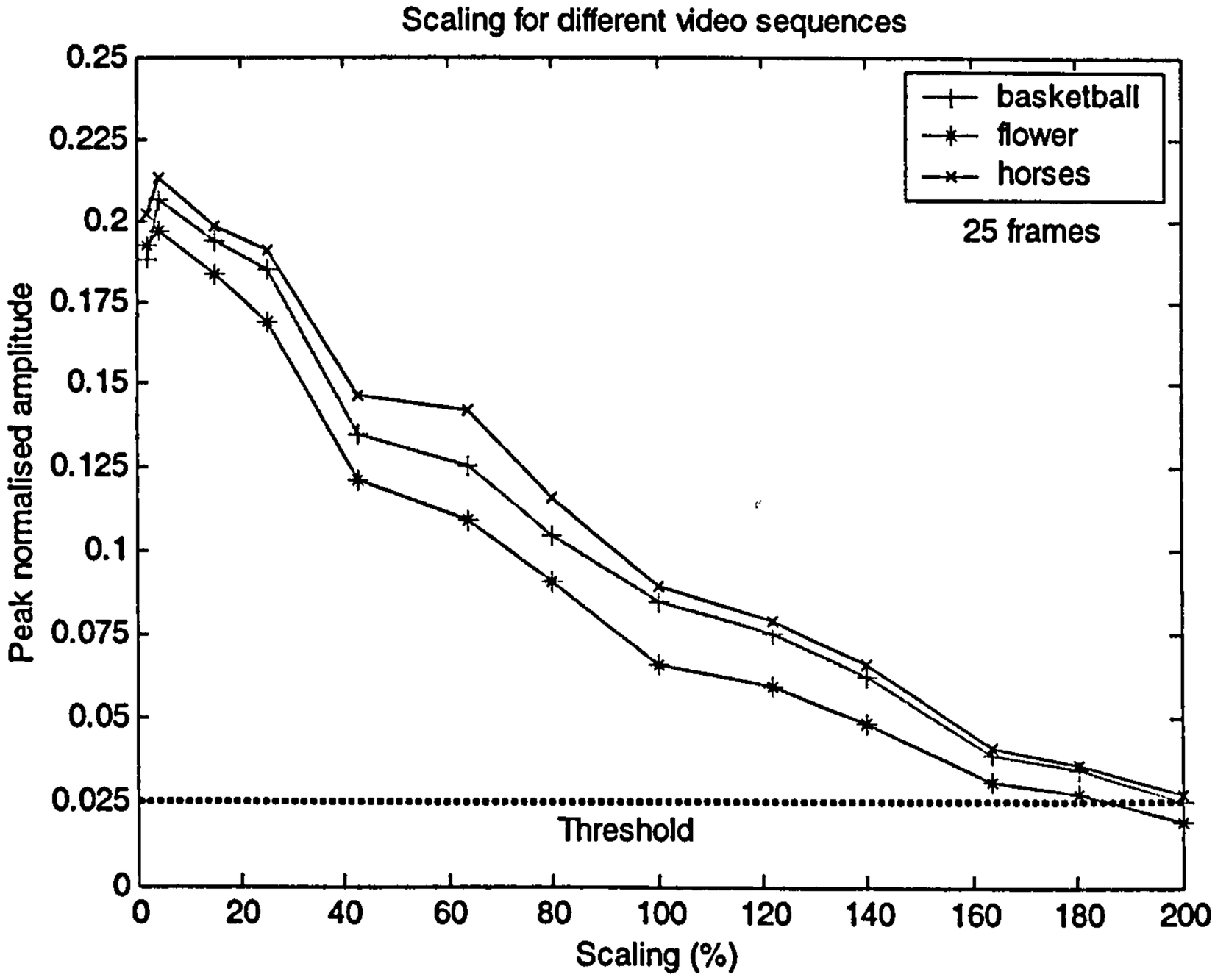


Figure 10. Peak normalised amplitude for different video sequences under scaling attack

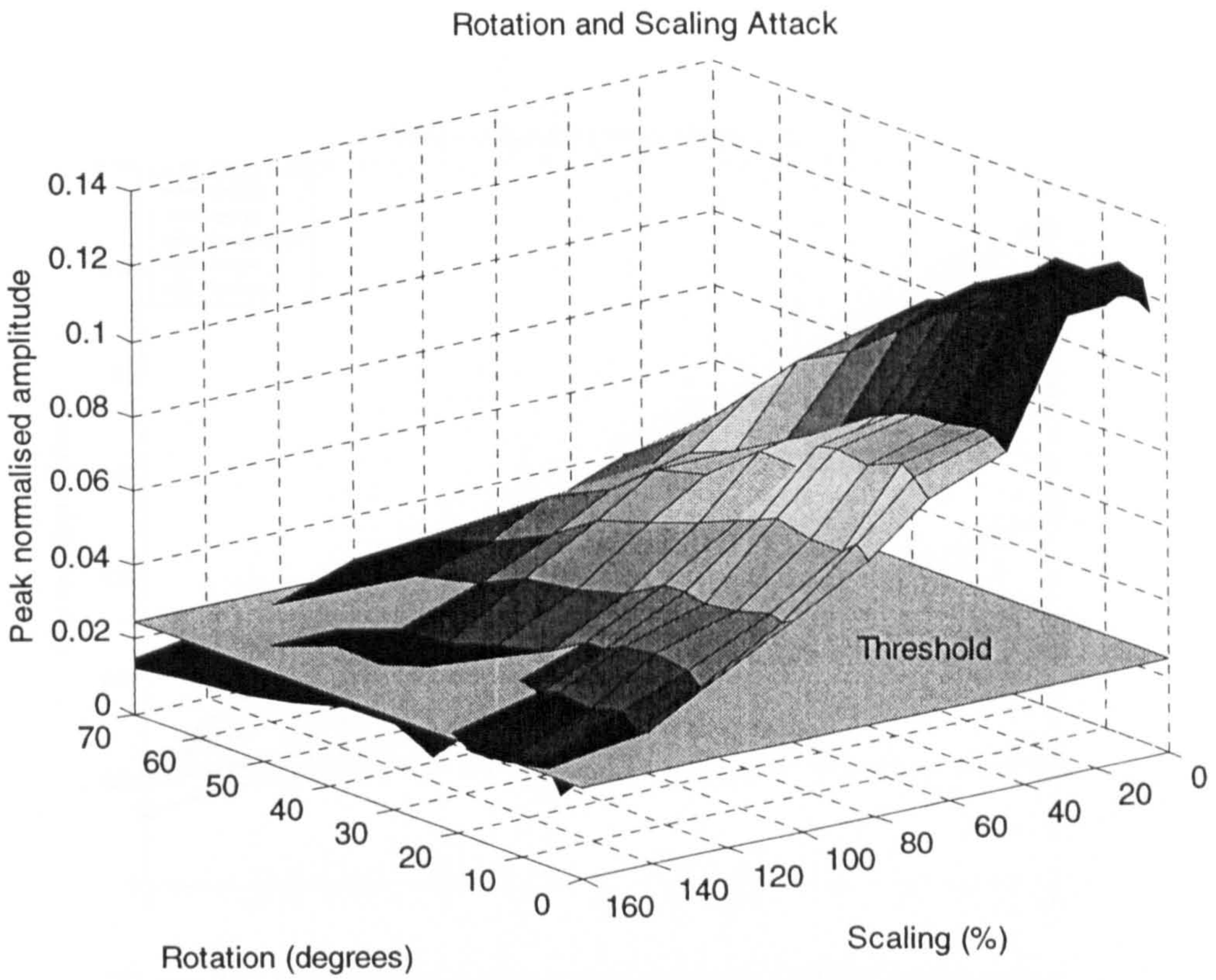


Figure 11. Performance of the system for rotation combined with scaling (25 frames)

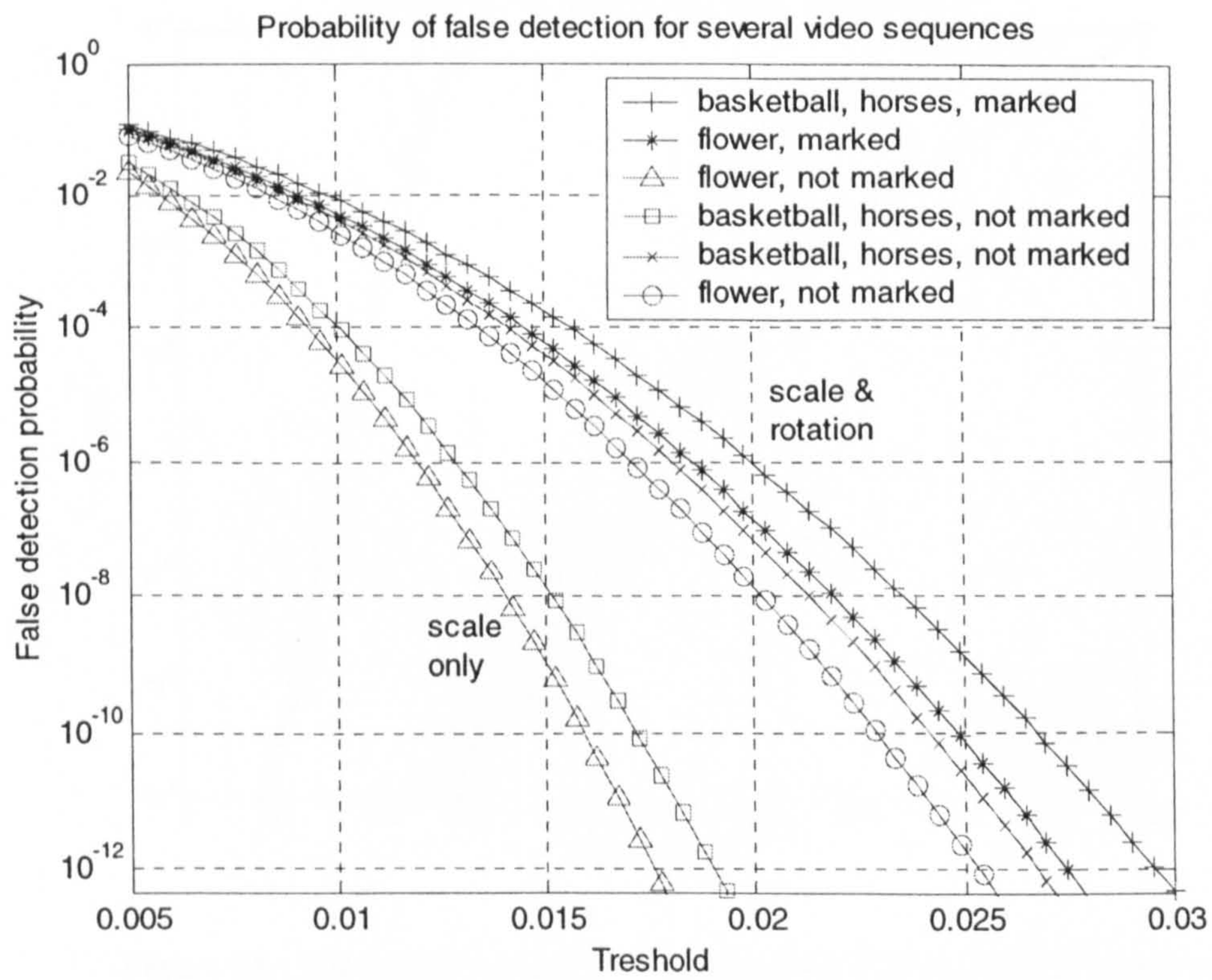


Figure 12. Threshold selection for a desired probability of false positive detection

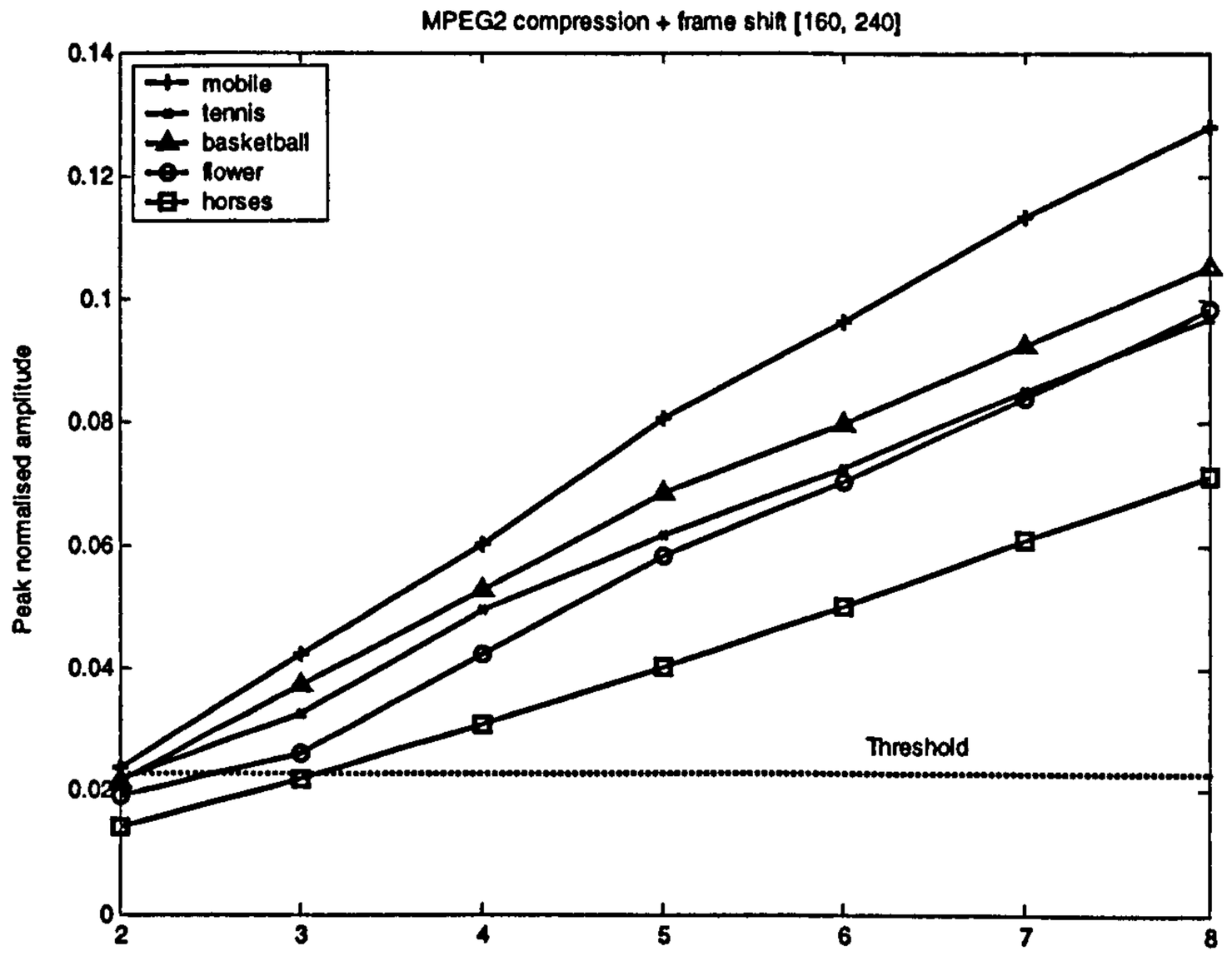


Figure 13. Performance under combined attack: MPEG2 compression combined with frame shift, for different video sequences (25 frames averaging).

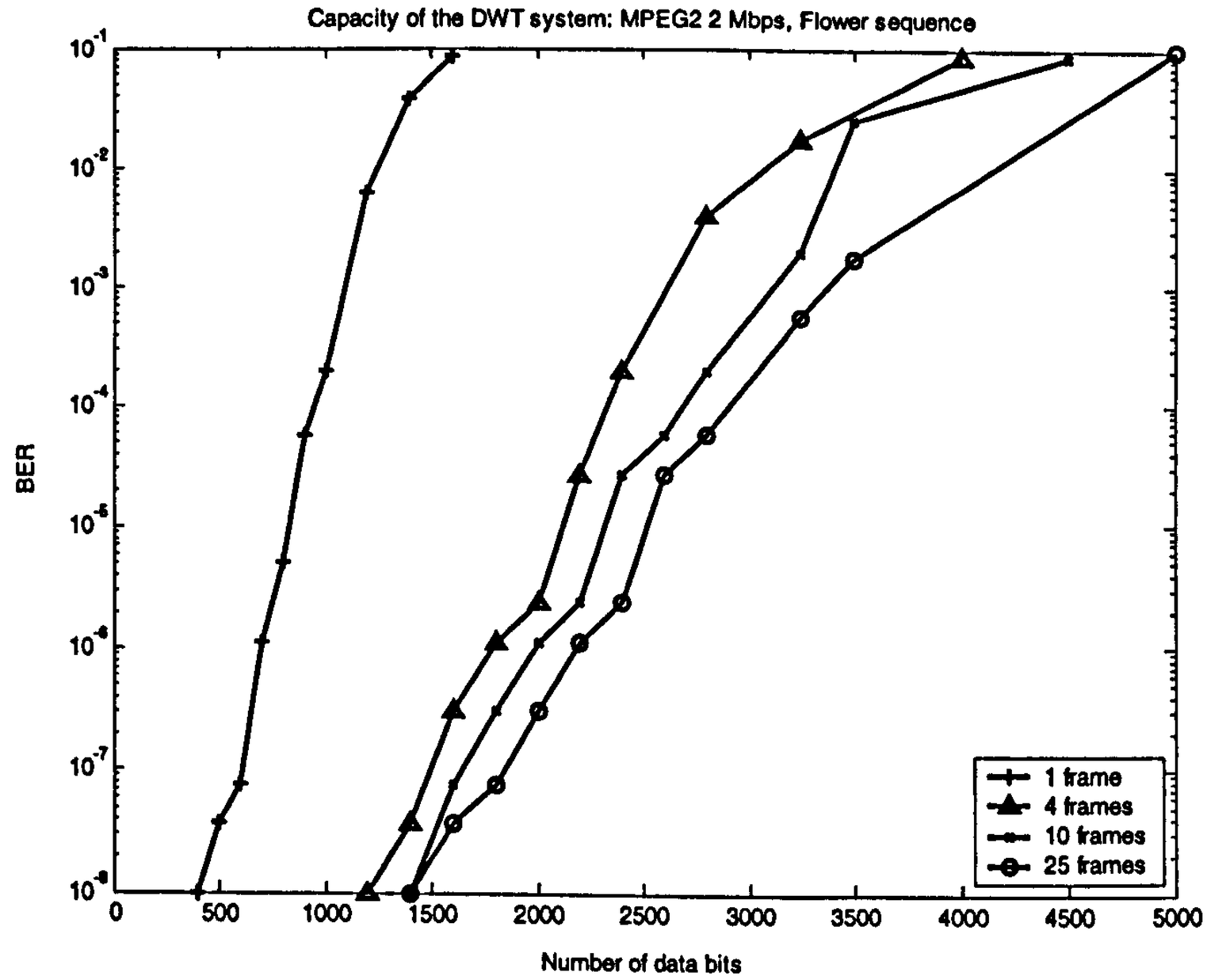


Figure 14. The capacity of the DWT system under 2Mbps MPEG2 compression attack when using frame averaging, for the "flower garden" video sequence.

Watermarking uncompressed video: an overview

Invited Paper

C. Serdean, A. Ambroze, and G. Wade (Department of Communication and Electronic Engineering, University of Plymouth)

M. Borda (Communications Department, Technical University of Cluj-Napoca, Romania)

I. Nafornita (Communications Department, Politechnica University of Timisoara, Romania)

Abstract

The paper overviews the watermarking of broadcast quality video, based upon the usual transform domain, spread spectrum approach. Aspects covered include perceptual-based marking, the advantages of wavelet based marking, and the performance of sliding correlators. The paper also derives an operational capacity for practical watermark channels, and explores the capacity improvement through the use of channel coding. The capacity is deduced by examining the correlation distribution at retrieval, and can be determined given MPEG-2 compression, geometric attack, visual thresholds, and channel coding. It is found that FEC based on multiple parallel concatenated convolutional codes (3PCCCs) can give over an order improvement in capacity for compressed video, and typically gives 0.5 kbit/s capacity even under a combined compression-geometric attack.

1 Introduction

Hidden data or a watermark is inserted into a video sequence for the purposes of copyright protection and video 'fingerprinting', and it can be performed either on uncompressed video (ITU-R 601) e.g. in studios, or on MPEG compressed video [1]. The former is discussed here, with emphasis upon the achievable channel capacity given typical attack conditions and advanced channel coding. The system under consideration is shown in Fig. 1. It uses the well-known spread spectrum approach, but also protects marking using FEC; N_b coded or uncoded bits are embedded over a sequence of frames, giving N_b normally distributed crosscorrelation peaks.

1.1 Requirements

Apart from the fundamental requirement that the watermark should not be visible under comparative studio viewing conditions, the design of a watermarking system for studios is subject to basic constraints [2], some of which are summarised in Table 1. Throughout this paper, we assume marking is on the luminance component of ITU-R 601 component digital signals in Y, Cr, Cb format (720×576 pixels).

1.	Watermark minimum segment (WMS) : 1 – 5 sec
2.	Bit error and false alarm probability: 10^{-8}
3.	Cryptographically strong, with a watermarking key
4.	Watermark detection: single-ended in that the unmarked host video is not available for retrieval (blind watermarking)
5.	Robustness to :
	• MPEG-2 compression (≥ 2 Mbit/s) • PAL encoding • analogue (VHS) and digital recording • Studio processing: A/D and D/A conversion, sampling rate/aspect ratio/frame rate conversion • Geometric attack: picture shift, cropping and scaling to minimum picture size, unnoticeable rotation • Collusion attack

Table 1 : Basic watermarking requirements for studio signals

A small WMS is required to facilitate studio editing (cut and paste). A minimum payload is 64 watermark bits/WMS, although here we will examine the potential for higher capacity. Attacks can be unintentional. For example, studio mixing can shift the frame 20 pixels, standards conversion can omit frames, and jitter can occur on VHS recording. The collusion attack is a serious problem; it averages multiple versions of the same video sequence (each version having a different fingerprint) with the objective of removing the fingerprint.

The requirements for video are different from those for image watermarking. For example, geometric attacks such as StirMark can perform serious rotation, scaling and translation (RST) on images and have led to the development of RST invariant transforms. For broadcast video this is less of a problem and the basic requirement here is robustness to *unnoticeable* rotation.

2 Watermark embedding

Watermark embedding can be carried out in either the spatial domain or the transform domain. Transform domain marking is generally preferred since it is easy to avoid marking high video frequencies (which tend to be attenuated by compression), and because it is naturally suited for perceptual marking based upon the human visual system (HVS). Also, from an information theoretic argument, transform domain marking can give increased channel capacity compared to spatial domain marking [3].

The FFT coefficients offer the possibility of phase modulation, but in order to obtain a real image/frame, it is necessary to maintain complex conjugate symmetry. Effectively this halves the potential marking capacity. In addition, the phase is quite sensitive to MPEG compression. Many transform domain schemes have been based upon modulation of the discrete cosine transform (DCT) coefficients, as shown in Fig.2. An advantage in developing a DCT-based system is that it is also appropriate for marking the MPEG-2 bitstream. However, recent work

on image watermarking has shown that marking the coefficients of the discrete wavelet transform (DWT) can have significant advantages, and this is discussed in section 5.2.

2.1 SNR in a video watermarking channel

Figure 2 shows the well known spread spectrum watermarking approach based upon the discrete cosine transform (DCT). For chip rate c_r, each data bit u_j is spread as $b_i = u_j,\ jc_r \le i < (j+1)c_r$, and the product $b_i p_i$ is formed, where $\{p_i\}$ is a binary $\{\pm 1\}$ PN sequence. This sequence spreads u_j over many 8×8 pixel blocks distributed over a number of video frames. Video dependent marking is an essential component of a successful marking scheme [4] if only that it ensures that marking energy is low in low detail areas of a video frame. A simple video dependent watermark is

$$w_i = \alpha b_i p_i |C_i| \tag{1}$$

Here, C_i is a DCT coefficient and α is selected to give a mark below the threshold of visual perception ($\alpha << 1$). If there is no attack or filtering ($\hat{C}_i' = C_i'$), and assuming $u_j = 1$, the normalized crosscorrelation is

$$d_j = \frac{1}{c_r} \sum_{i=jc_r}^{(j+1)c_r-1} p_i C_i + \frac{\alpha}{c_r} \sum_{i=jc_r}^{(j+1)c_r-1} |C_i| p_i^2 \tag{2}$$

We are interested in the distribution of these crosscorrelation peaks since it determines the detected bit error rate (BER). If, for simplicity, we assume C_i is i.i.d. with $C_i \sim N(0, \sigma_c^2)$, then d_j has the approximate distribution

$$d_j \sim N\left(\alpha \mu_{|C|}, \frac{\sigma_c^2}{c_r}\right) \tag{3}$$

where $\mu_{|C|}$ is the mean of $|C_i|$, $\alpha << 1$ and $p_i \in \{\pm 1\}$. The normal distribution follows from the Central Limit theorem since the cross correlator performs a sequence of correlation sums. Clearly, the BER will reduce as c_r increases (as expected) due to reduced distribution variance. The variance of d_j arises mainly from the first term in (2) and so we conclude that the non-zero crosscorrelation of the PN sequence with the DCT coefficients is a source of noise in the channel.

In practice, the distribution will be significantly affected by other factors. For example, it is widely recognised that crosscorrelation can be improved by inserting a 3×3 spatial filter in the video path (Fig.2). This removes low frequency video components prior to crosscorrelation and gives a distribution with larger mean and smaller variance. In practice, compared to filtering, balancing the PN sequence to ensure that it has zero mean gives only a relatively small improvement. Since marking is video dependent, the distribution will also depend upon the choice of sequence. Figure 3 illustrates the point for two standard MPEG video test sequences (ITU-R 601 format); it is apparent that sequence 'flower garden' will have a larger BER than sequence 'mobile'. The underlying normal distribution is shown dotted.

Generalising, for any particular system the distribution mean μ, and variance σ^2 define a SNR of the channel: $SNR = (\mu/\sigma)^2$. The corresponding BER for an uncoded system is simply $BER_u = Q[\mu/\sigma] = Q[\sqrt{SNR_u}]$. For a coded system, μ and σ define a signal to noise ratio SNR_c at the decoder input, and the decoded bit error rate is $BER_c = f(SNR_c)$ where f is a known function for a particular iterative decoder (Fig.6(b)).

2.2 Perceptual-based marking

Rather than use (1), a better approach is to base marking on both the video sequence and the HVS. In this paper we compute a perceptual threshold or *just noticeable difference*, JND_i, for each DCT coefficient [5][6] and watermark as

$$C_i' = C_i + \alpha b_i p_i JND_i \tag{4}$$

where α is a small constant (see Fig.2). Marking is HVS based since JND_i is computed from physcovisual properties of the eye, and it is video dependent since JND_i is also a function of C_i. Human perception is also incorporated by making JND_i a function of the MPEG-2 default quantization matrix elements Q_i [7].

We compute *JND* values using a simplified form of Watson's model [8] and then enhance it to account for lateral inhibition masking [9]. First we compute the frequency sensitivity (modulation transfer function) of the eye as $T_F(i) = Q_i/2$. This greatly simplifies Watson's approach and avoids the need for empirical parameters. The *luminance masking* thresholds T_L and *contrast masking* thresholds T_C for block k are then [8]

$$T_L(i,k) = T_F(i) \cdot \left[\frac{C_{0,k}}{\overline{C}_0}\right]^{0.649} \tag{5}$$

$$T_C(i,k) = T_L(i,k) \cdot \max\left[1, \left(\frac{|C_{i,k}|}{T_L(i,k)}\right)^{w}\right] \tag{6}$$

where $\overline{C}_0$ corresponds to the mean DC coefficient over a frame, and $w = 0$ for the DC coefficient and $w = 0.7$ elsewhere. Equation (6) accounts for three aspects of the HVS, and, in a similar way to [9], it could be extended by incorporating lateral inhibition masking:

$$T_{LI}(i,k) = \left\{ \begin{array}{l} T_C(i,k) \quad \text{if} \left(T_C(i,k) > \mu(N_{i,k}) \text{ or } (T_C(i,k) - \sigma(N_{i,k})) < \dfrac{Q_i}{32} \right) \\ T_C(i,k) - \sigma(N_{i,k}) \quad \text{otherwise} \end{array} \right\} \tag{7}$$

where $\sigma(N_{i,k})$ and $\mu(N_{i,k})$ are the standard deviation and the mean for the eight neighbours of $T_C(i,k)$. We obtained the condition in (7) from subjective tests. Using the basic *JND* definition in [8], the values for block k are then

$$JND_{i,k} = \frac{Q_i}{2T_{LI}(i,k)} \tag{8}$$

Equation (8) ensures that coefficients that are less visible are marked with greater energy. In practice, the theoretical *JND* values have been found to be within a factor 2 or 3 of the actual perceptual threshold, and this is accounted for by the factor α in (4).

3 *Channel capacity*

If we regard the watermark channel as a communications system with input X (the watermark data) and output Y, the channel capacity is defined as the maximum mutual information:

$$C_{chan} = \max_{p(x)} I(X;Y) = \max_{p(x)}[h(X) - h(X \mid Y)] = \max_{p(x)}[h(Y) - h(Y \mid X)] \qquad (9)$$

where the maximum is taken over all possible distributions $p(x)$. Term $h(X \mid Y)$ represents information loss due to channel 'noise', which will be a combination of the host video and signal processing (compression/attack). If the loss is modelled as the addition of an independent Gaussian noise source, $Z \sim N(0,\sigma_z^2)$, i.e. $Y_i = X_i + Z_i$, where Z is a continuous random variable, then (9) reduces to [10]

$$C_{chan} = \frac{1}{2}\log_2\left(1+\frac{\sigma_x^2}{\sigma_z^2}\right) \text{ bits/symbol} \qquad (10)$$

providing $X \sim N(0,\sigma_x^2)$. In [3], σ_x^2 was estimated from acceptable JPEG performance, and an equivalent Gaussian noise variance was computed for the host and JPEG compression. In [11] the noise was restricted to an AWGN attack, the host noise being virtually eliminated by using it as side information during embedding.

In this paper we invoke *JNDs* to maximise the signal power. From the discussion in section 2.1, the channel in Fig.1 can be modelled as a gain factor cascaded with a Gaussian noise source $Z \sim N(0,\sigma_z^2)$ (the gain and variance depending upon the host video, MPEG compression, and geometric attack). For example, we could estimate a basic operational capacity as follows. Suppose that all N_p pixels ($N_p = 720\times576$) in the frame are transformed via DCT blocks, and that the channel noise is simply that of the host video. Assuming C_i is i.i.d. for simplicity, the noise power per video frame is $N_p\overline{C^2}$, where $\overline{C^2}$ is the mean coefficient power for a particular video sequence. If only one data bit is embedded per video frame (corresponding to $c_r = N_p$), and there is no FEC, the SNR is

$$SNR = N_p\frac{\left(\alpha\,\overline{JND}\right)^2}{\overline{C^2}} = N_p\,\overline{SNR} \qquad (11)$$

where $\overline{SNR}$ is a measured mean SNR for the video sequence and $\overline{JND}$ is the mean JND for the sequence. If N_b data bits are embedded into a frame the signal to noise ratio per uncoded data bit reduces to $SNR_u = SNR/N_b$, and the data rate or capacity for an uncoded system of frame rate F_r is

$$D_r = N_b F_r = \frac{k\,N_p\,\overline{SNR}\,F_r}{SNR_u} \text{ bits/s} \qquad (12)$$

In (12) we account for the fact that, in general, only a fraction k of the coefficients in each DCT block are marked. Also, in the following work we will define the operational capacity as the maximum value of D_r for which the *BER* does not exceed a tolerable level (typically 10^{-8}).

Equation (12) has been used to estimate the capacity for the uncoded, uncompressed video test sequence 'flower garden' (graph (b) in Fig. 4), assuming the *JND*-based marking in (4). Coefficient α in (4) was set to give marking below the threshold of visibility (visible *JND* marking appearing as fine grain noise). Graph (a) in Fig.4 shows the corresponding simulated capacity, obtained by embedding N_b bits over N_f frames ($D_r = N_b F_r / N_f$) using (4). This gives N_b correlation peaks and the resulting distribution gives $SNR_u = (\mu/\sigma)^2$. The discrepancy between graphs (a) and (b) is attributed to highpass filtering prior to crosscorrelation. Graph (c) shows the simulated result for MPEG-2 compression to 6 Mbit/s. The reduced capacity due to compression is more clearly shown in Fig.5. Using $BER_u = Q\left(\sqrt{SNR_u}\right)$, graph (a) shows that the uncompressed capacity is around 3 kbit/s, and graph (b) gives experimental confirmation of (a) by directly counting data errors. Graph (c) shows that MPEG-2 compression reduces the capacity to the order of 100 bit/s.

4 Use of Turbo codes

Consider a coded system of rate R (Fig.1) and assume that N_b coded bits are embedded over a sequence of N_f video frames. The watermark data rate is now

$$D_r = \left(\frac{N_b R}{N_f}\right) F_r \text{ bits/sec} \tag{13}$$

Assuming as before that all N_p pixels in a frame are transformed, the spread spectrum chip rate is

$$c_r = \frac{k N_f N_p}{N_b} = \frac{k N_p R}{D_r} \cdot F_r \tag{14}$$

According to (14), for a fixed D_r, the use of FEC reduces the chip rate by a factor R. As indicated in (3), this increases the variance of the channel distribution, resulting in increased *BER*, and the FEC decoder must more than compensate for this increase in order to provide coding gain.

For a channel with a potentially large BER (due to attack) it is essential to use soft decision decoding, and in practice this restricts the choice of FEC to convolutional codes. Viterbi decoding is the usual ML decoder for an AWGN channel, and so has been used to protect watermarked video [12]. Here, we are interested in the recovery of channel capacity that can be achieved by using Turbo codes in an attacked watermarking channel. Turbo codes offer better coding gain and could be used at low rate in this application (with consequent improvement in performance). We consider a code of rate R and interleaver size N, and treat the FEC as block coding (block length N). Note from Fig.1 that coding precedes spread-spectrum so that the input to the iterative decoder will be the output of the spread-spectrum correlator. This scheme ensures that the watermark channel up to the decoder input approximates to a Gaussian channel, and the latter is the usual assumption for Turbo code

systems. An alternative scheme is to place the FEC encoder after the spread-spectrum process. This has the potential to provide large block lengths for the encoder and so improve its performance, but has the disadvantage that the channel at the decoder input is poorly defined and will have a very low SNR.

In this paper a multiple parallel concatenated convolutional code (3PCCC) has been used to protect the watermark channel and the encoder is shown in Fig.6(a). The use of two interleavers (I_1 and I_2) rather than one as in the basic Turbo code reduces the error rate floor and so gives improved performance [13]. Each recursive systematic code (RSC) is an optimum (5,7) code [14], giving an unpunctured code rate $R = 1/4$. Figure 6(b) shows the simulated performance of the overall code for several interleavers.

Figure 4 shows that an uncoded watermark channel at 1 kbit/s corresponds to $SNR_u \approx 4$ dB when 6 Mbit/s MPEG-2 compression is used. Whilst this low SNR is unusable without FEC, Fig.6(b) shows that that iterative decoding should be effective against this sort of attack. Figure 7 illustrates the effectiveness of iterative decoding against a 6 Mbit/s MPEG-2 attack, and also compares the performance of the heuristic marking scheme in (1) (Fig.7(a)) with that of the *JND* scheme in (4) (Fig.7(b)). The BER of the uncoded system is computed as before, whilst that of the coded system is computed as $BER_c = f(SNR_c)$ using Fig.6(b). Without FEC the attack reduces the 10^{-8} capacity to around 300 bits/s, but with FEC the capacity can be over 3 kbit/s. In each case, α was selected to give marking just below the threshold of visibility. Note that *JND* marking is superior to the heuristic scheme, and can give a capacity over 8 kbit/s for uncompressed video (not shown).

4.1 Combined attacks

In practice watermarked video is likely to suffer from a combination of attacks, such as MPEG-2 compression and geometric distortion, and an attack of this nature can defeat many watermarking schemes [15][16]. Figure 8 shows the effect of a combined compression and line cut attack upon the *JND*-based marking scheme in (4). In order to combat the line cut we use a 2-D sliding correlator. This moves the known PN sequence over a small search window relative to the received coefficient block in order to locate the correlation peak. The 10^{-8} capacity is relatively low for an uncoded system (graph (a)), and a possible explanation is that compression reduces the watermark amplitude to below the performance threshold of the sliding correlator. However, graphs (b) and (c) show that performance can be significantly improved through the use of FEC.

5 System improvements

5.1 3-D correlator

Spread spectrum systems are vulnerable to synchronisation error, as can occur in a geometric attack. In Fig.8 we used a 2-D (spatial) correlator to combat a line cut attack, and Fig.9 shows a 3-D (spatial/temporal) correlator for combating temporal attacks, such as a frame cut. All three macro blocks shown are marked with the same data bit, but they are randomly placed within frames (as indicated) in order to improve security (within a macro block the data is at the same spatial location in order to perform correlation, the typical marking depth being four frames).

The use of a sliding correlator has the disadvantage of decreasing the effective chip rate since a local crosscorrelation peak is computed for small blocks (the overall crosscorrelation being the sum of the local correlations). This amounts to a 'correlator loss', and is illustrated in Fig.10. The loss is about 3 dB for the 3-D correlator (Fig.10 graph (b)), and is the same for a frame cut (irrespective of position within the sequence). If a 2-D correlator is used, the loss varies significantly and can be large e.g. 12 dB if the frame cut occurs at frame 30 within a 120 frame sequence (Fig.10 graph (d)).

5.2 Use of the DWT

The human observer tends to process video information by independently processing multiple frequency channels. In a similar way, the multiresolution decomposition of the discrete wavelet transform (DWT) also enables signals in different channels to be processed independently. *This implies that frequency bands can be processed independently without significant perceptible interaction between them, and that the need for a complex perceptual marking algorithm (JNDs) is reduced.* Looked at another way, the multiresolution aspect also enables a watermark to have spatially local and spatially global components.

Figure 11(a) illustrates a simple watermarking scheme based on the DWT [17]. For one level of decomposition, the DWT generates 4 subimages (coefficients) gg, gh, hg, hh. A second level of decomposition is obtained by similarly transforming the approximation coefficients gg. Watermarking is performed by multiplying the *detail* coefficients gh, hg, hh by a (secret) constant, K, its value being chosen to avoid visible artefacts. This procedure tends to mark perceptually significant regions in a video frame e.g. edges, as required for robust watermarking. The watermarked frame is I_w and the watermark itself is W. Assuming no attack for simplicity, Fig. 11(b) shows that the watermark can be extracted blind by multiplying the detail coefficients gh, hg, hh by K^{-1}. In the presence of an attack, $W' \neq W$ and watermark detection can be carried out by crosscorrelating W' with W, and thresholding the result. Clearly, this simple system conveys just 1 bit of information. Also, more secure marking would use a PN sequence to identify coefficients to be marked.

Using the system in Fig.11, it has been shown that DWT marking is more robust to compression compared to DCT marking [17]. More comprehensive work confirms this and shows that significant improvements can also obtained wrt geometric attacks e.g. frame shift, cropping and scaling [18]. The watermarking component with local spatial support tends to be robust to cropping, whilst the component with global support tends to be robust to lowpass filtering.

5.3 Choice of PN sequence

In order to provide a degree of security, each data bit in Fig.2 is spread using a different PN sequence. Each PN sequence is a subset of a large PN sequence, and the key to this sequence is also used to select random locations for the marked blocks. Evenso, the multiplicative congruential algorithm [19] used to generate the large sequence is known to be insecure. In general, the PN sequences used should have the following properties:

- large chip rate e.g. $c_r \approx 10^6$
- large autocorrelation and small crosscorrelation

- cryptographically secure

If we use selected pairs of m sequences (*preferred sequences*), it is possible to obtain good autocorrelation and reasonably low cross-correlation. For an n stage LFSR the cross-correlation function can then be 3-valued: $\{-1, -t, t-2\}$ where $t = 2^{\lfloor (n+2)/2 \rfloor} + 1$. Better cross-correlation is achieved using Kasami sequences, the 3-values being: $\{-1, -(2^{n/2}+1), 2^{n/2}-1\}$, n even. For $n=6$, preferred pairs would give $\{-1, -17, 15\}$, compared to $\{-1, -9, 7\}$ for Kasami sequences. However, an m sequence generated from an n-stage linear feedback shift register is not cryptographically secure since it can be deduced from just $2n$ bits of the sequence. A partial but still unsatisfactory solution is to associate some non-linearity with the register. In order to improve security, Kasami sequences could be encrypted using one of the well known algorithms e.g. DES ECB, DES CBC, IDEA ECB, or IDEA CBC [20]. Experiment shows that the resulting crosscorrelation is somewhat inferior to pure Kasami sequences [17].

6 Conclusions

A spread spectrum-based video watermark data channel is conveniently characterised by the Gaussian distribution at the output of the sliding correlator. This distribution defines a SNR for the channel, from which can be deduced an operational channel capacity for a system subject to perceptual marking, combined attack, and channel coding. The Gaussian input to the FEC decoder, and the fact that low code rates can be tolerated, makes iterative decoding particularly appropriate for the protection of a watermarked channel. The computational complexity of such decoding is still relatively small compared to that of the sliding correlator.

As expected, channel capacity increases through the use of perceptual marking (*JNDs*) and FEC, and reduces when watermarked video is subjected to attacks. MPEG-2 compression to 6 Mbit/s reduces the 10^{-8} capacity from over 8 kbit/s (uncompressed video, *JND* marking) to about 300 bits/s, although an order improvement is achieved through FEC. The vulnerability of spread spectrum systems to synchronisation error has been highlighted, although a 3-D correlator has proved effective against temporal attack, such as a frame cut. A combination of MPEG-2 compression and simple geometric attack can severely reduce capacity, although useful improvements can still be made through the use of FEC. Under such an attack, FEC enables the current watermarking scheme to achieve a typical capacity of 500 bits/s, although this is very video dependent. Increased robustness to attack should be achievable by replacing the DCT with the DWT.

References

1. Hartung, F., and Girod, B. 'Digital Watermarking of Raw and Compressed Video', *Proc. SPIE 2952: Digital Compression Technologies and Systems for Video Communication'*, October 1996, 205-213.

2. Union Europeenne de Radio-Television (EBU-UER) Watermarking Working Group, February 2000.

3. Ramkumar, M., and Akansu, A. 'Information Theoretic Bounds for Data Hiding in Compressed Images', IEEE 2nd Workshop on Multimedia Signal Processing, Redondo Beach, CA, 1998.

4. Swanson, M., Zhu, B., Chau, B., and Tewfik, A. 'Object-based transparent video watermarking', *Electronic Proc. IEEE Signal Processing Society*, 1997 Workshop on Multimedia Signal Processing', Princeton, New Jersey, 23-25 June, 1997.

5. Wolfgang, R., Podilchuk, C., and Delp, E. 'Perceptual Watermarks for Digital Images and Video', *Proc. IEEE*, Vol. 87, No.7, July, 1999, 1108-1126.

6. Kim, S., Suthaharan, S., Lee, H., and Rao, K. 'Image watermarking scheme using visual model and BN distribution', *Electronic Letts.*, Vol. 35, No.3, February 1999, 212-213.

7. ISO/IEC 13818-2, 'Information Technology - Generic coding of moving pictures and associated audio information'.

8. Watson, A. 'DCT quantization matrices visually optimised for individual images', Human Vision, Visual Processing and Digital Display IV', *Proc. SPIE 1913-14*, 1993, 1-15.

9. Kim, S. 'Image watermarking scheme using visual model and BN distribution', *Electronics Letts.*, Vol. 35, No.3, 4 Feb. 1999, 212-213.

10. Cover, T., and Thomas, J. 'Elements of Information Theory', Wiley, 1991.

11. Eggers, J., Su, J., and Girod, B. 'A Blind Watermarking Scheme Based on Structured Codebooks', IEE Secure Images and Image Authentication, London, UK, April 2000, 1-6.

12. Hernandez, J., Delaigle, J., and Macq, B. 'Improved data hiding by using convolutional codes and soft-decision decoding', *Proc. SPIE*, 'Security and Watermarking of Multimedia Contents', San Jose, CAL., 24-26 January 2000, 24-47.

13. Divsalar, D. and Pollara, F. 'Multiple Turbo Codes for Deep Space Communications', *TDA Progress Report 42-121*, May 1995, 66-77.

14. Benedetto, S., Garello, R. and Montorsi, G. 'A Search for Good Convolutional Codes to be Used in the Construction of Turbo Codes', *IEEE Trans. on Communications*, Vol.46, No.9, Sept. 1998, 1101-1105.

15. Petitcolas, F., Anderson, R., and Kuhn, M. 'Attacks on Copyright Marking Schemes', Second Workshop on Information Hiding', in Vol. 1525 of *Lecture Notes in Computer Science*, Portland, Oregon, USA, 218-238, 14-17 April, 1998.

16. Hartung, F., Su, J., and Girod, B. 'Spread Spectrum Watermarking: Malicious Attacks and Counterattacks', *Proc. of SPIE*, Vol. 3657: Security and Watermarking of Multimedia Contents, January 1999.

17. 'Preliminary Report on Watermarking', Borda, M. and Deac, V., Communication Dept., Technical University of Cluj-Napoca, Romania, and Nafornita, I. and Isar, A., Politechnica University of Timisoara, Romania, June 2000.

18. Podilchuck, C., and Zeng, W. 'Image-Adaptive Watermarking Using Visual Models', IEEE Journal, SAC, Vol.16, No.4, May 1998, 525-538.

19. Press, W. and Teukolski, S. 'Numerical recipes in C', Cambridge University Press, 1993.

20. Stallings, W. 'Network and Internetwork Security Principles and Practice', Prentice Hall, 1995.

Figure Captions

Fig.1: Channel coding in a watermark channel.

Fig.2: Transform domain spread spectrum watermarking and retrieval.

Fig.3: Channel distributions for two video sequences ($u_j = 1 \ \forall j$)

Fig.4: SNR for sequence 'flower garden' ; (a) uncompressed, filtered; (b) uncompressed, unfiltered, equation (12); (c) MPEG-2 compressed to 6 Mbit/s, filtered.

Fig.5: BER for sequence 'flower garden'; (a) uncompressed; (b) uncompressed simulation; (c) MPEG-2 compressed to 6 Mbit/s, filtered.

Fig.6: Rate ¼ 3PCCC FEC: (a) encoder; (b) simulated performance of an iterative decoder for interleaver sizes of 500 and 2000.

Fig.7: Combating a 6 Mbit/s MPEG-2 attack with iterative decoding ($N = 2000$) for sequence 'basketball': (a) heuristic marking, $\alpha = 0.004$; (b) JND marking, $\alpha = 3$.

Fig.8: Combined compression and line cut attack on sequence 'basketball': (a) uncoded; (b) coded, $N = 500$; (c) coded, $N = 2000$; (d) Shannon limit.

Fig.9: 3-D sliding correlator.

Fig.10: Correlator loss for sequence 'flower garden' (120 frames) : (a) no attack, no sliding; (b) 3-D correlator with/without frame cut attack; (c) 2-D correlator, frame cut attack after 60 frames; (d) 2-D correlator, frame cut attack after 30 frames.

Fig. 11: DWT domain watermarking: (a) embedding; (b) retrieval, assuming no attack

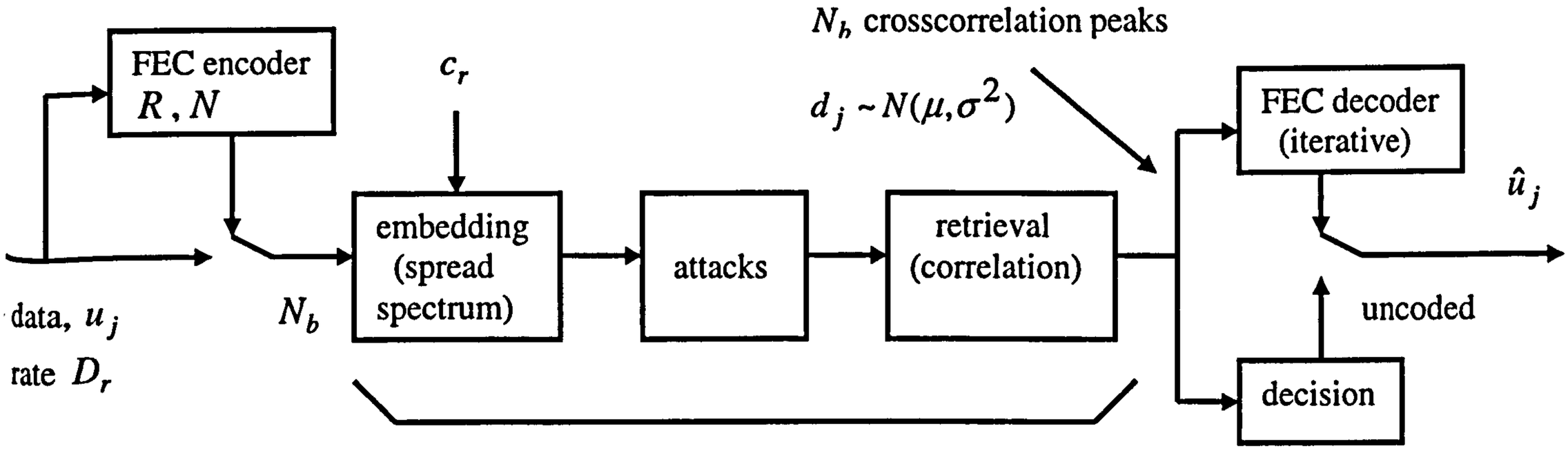


Figure 1 : channel coding in a watermark channel

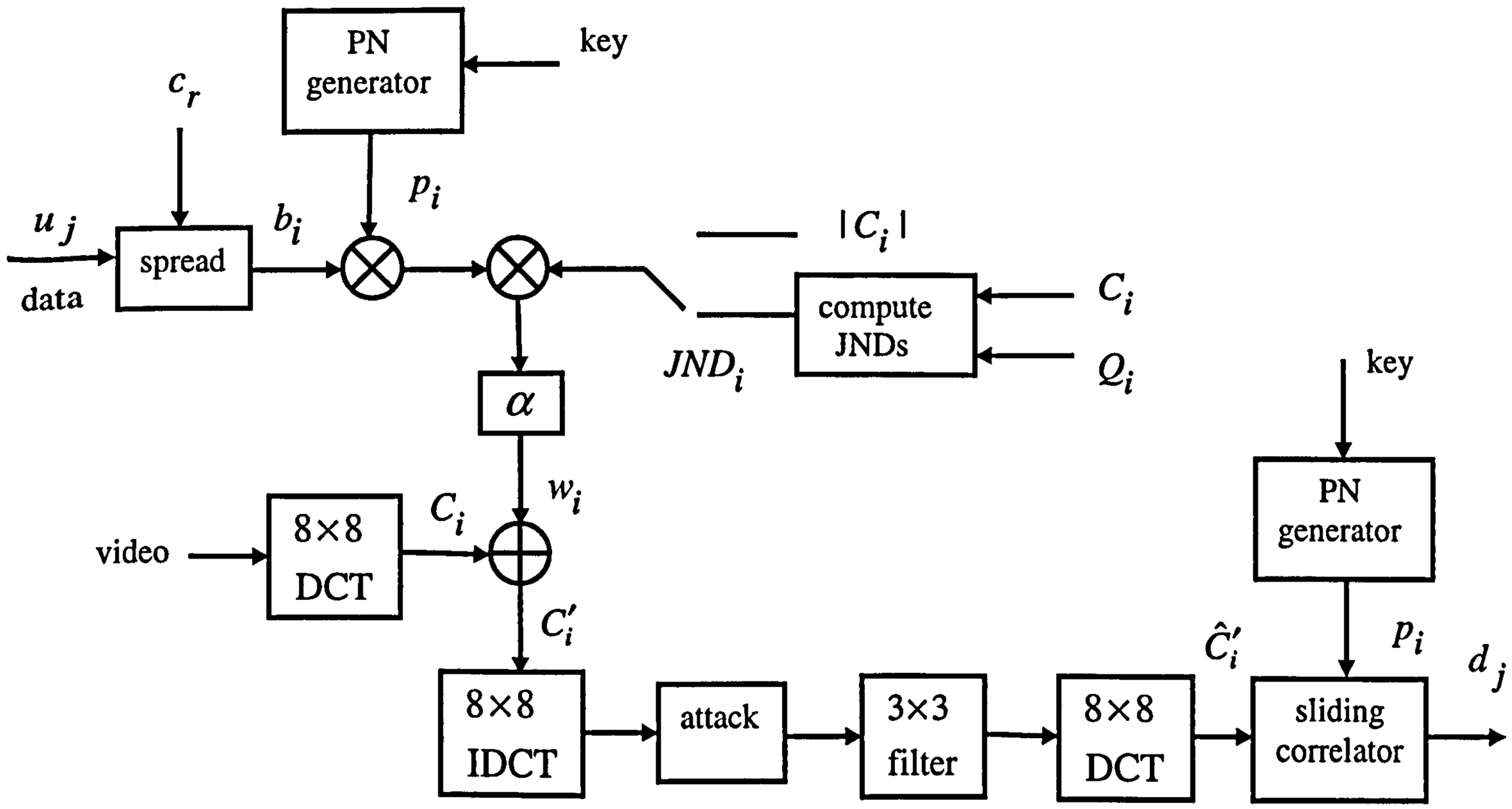


Figure 2 : Transform domain spread spectrum watermarking and retrieval

Fig 3

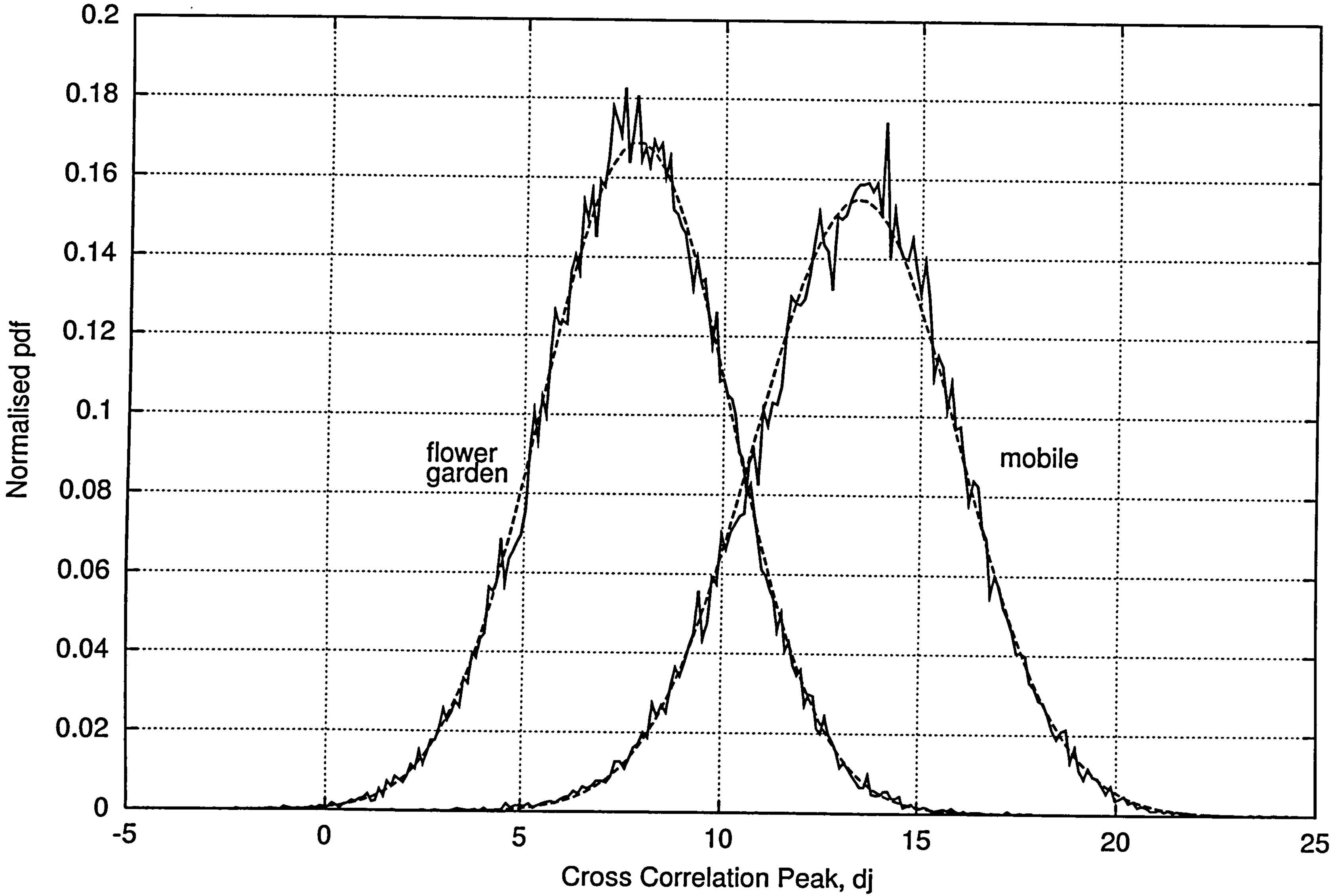

0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
Normalised pdf
flower garden
mobile
-5
0
5
10
15
20
25
Cross Correlation Peak, dj

Fig 4

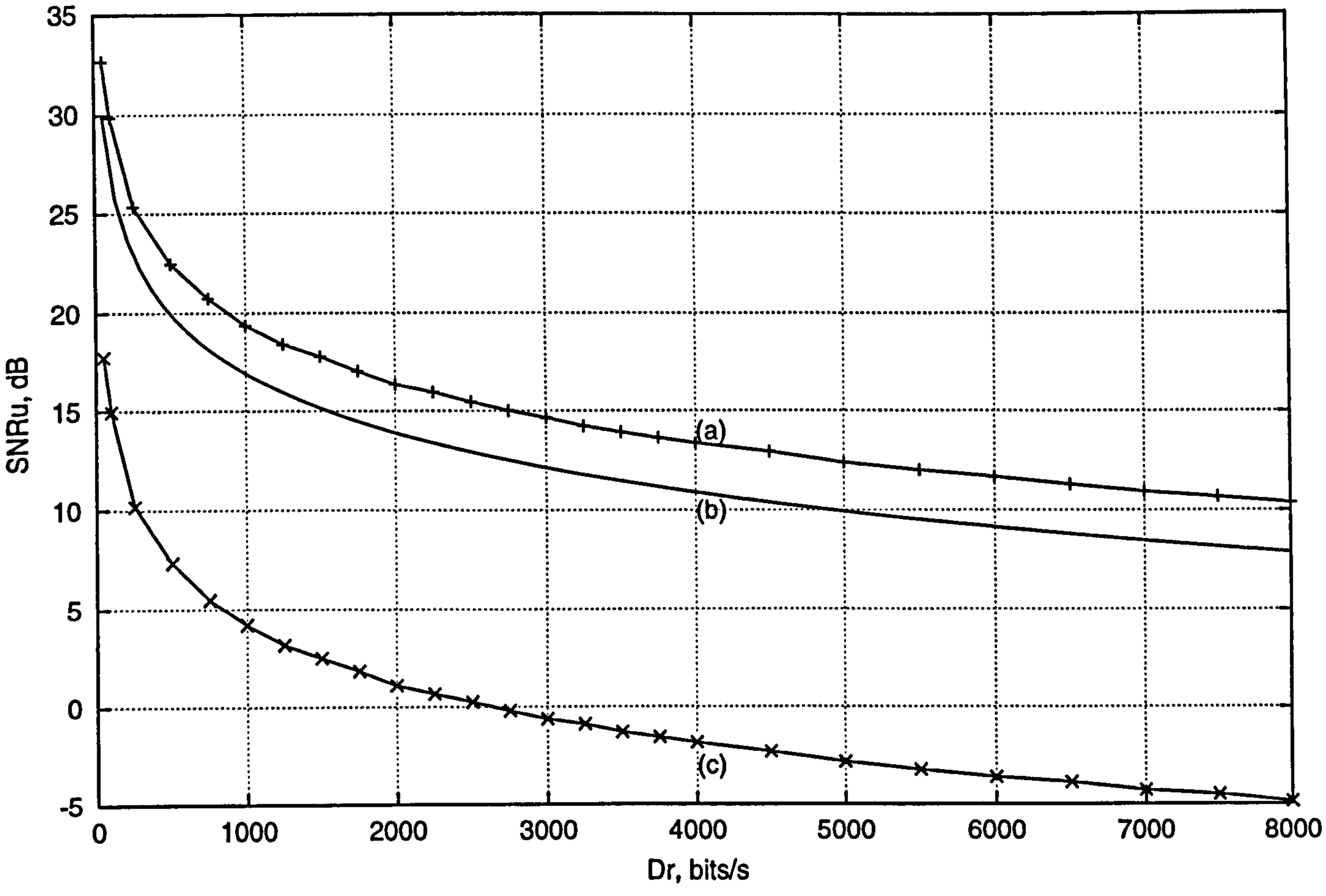
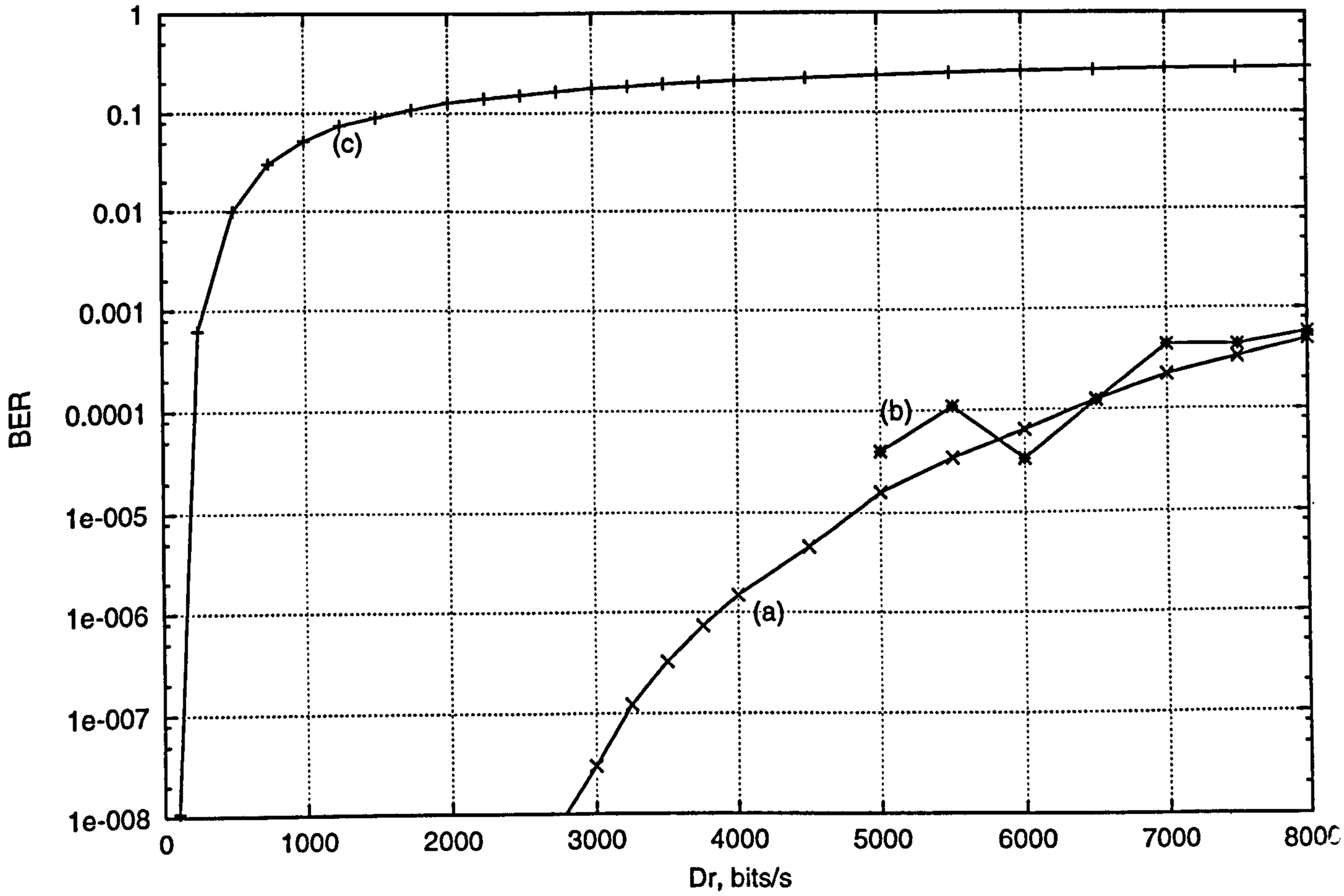

35
30
25
20
15
10
5
0
-5
SNRu, dB
(a)
(b)
(c)
0
1000
2000
3000
4000
5000
6000
7000
8000
Dr, bits/s

Fig 5


1
0.1
0.01
0.001
0.0001
1e-005
1e-006
1e-007
1e-008
BER
(c)
(b)
(a)
0
1000
2000
3000
4000
5000
6000
7000
8000
Dr, bits/s

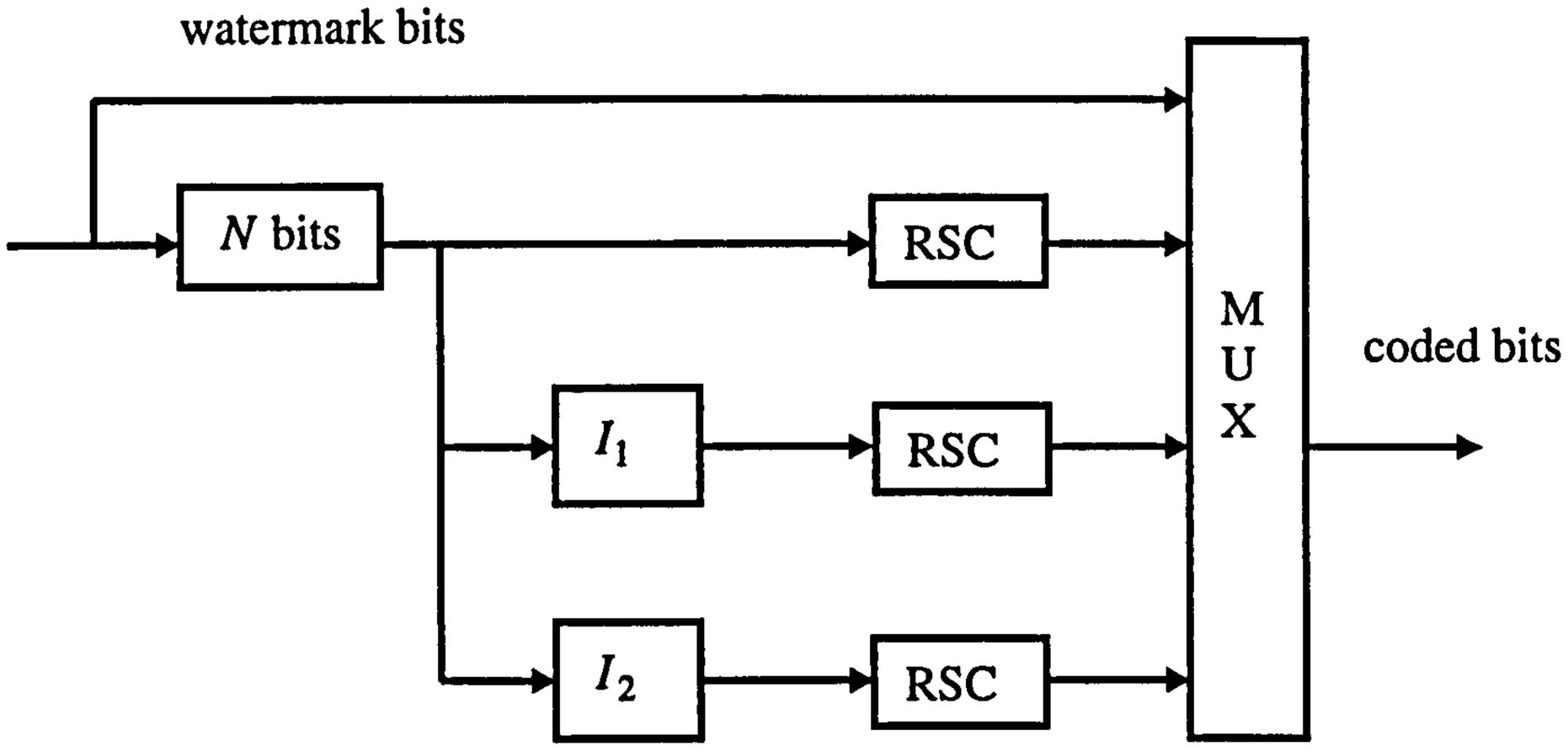


Fig. 6(a) Rate ¼ 3PCCC FEC encoder

Fig 6(b)

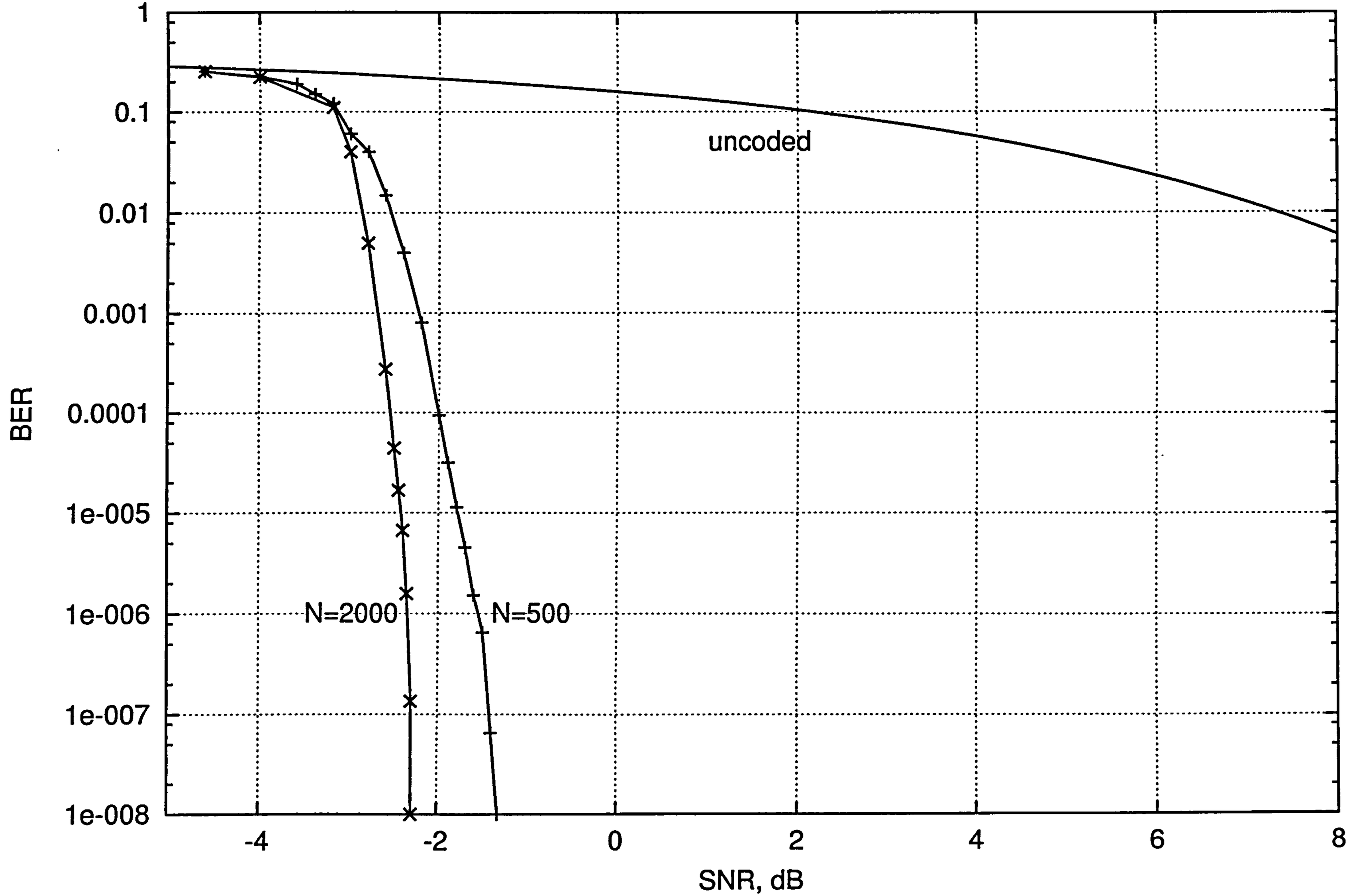

1
0.1
0.01
0.001
0.0001
1e-005
1e-006
1e-007
1e-008
BER
uncoded
N=2000
N=500
-4
-2
0
2
4
6
8
SNR, dB

Fig 7(a)

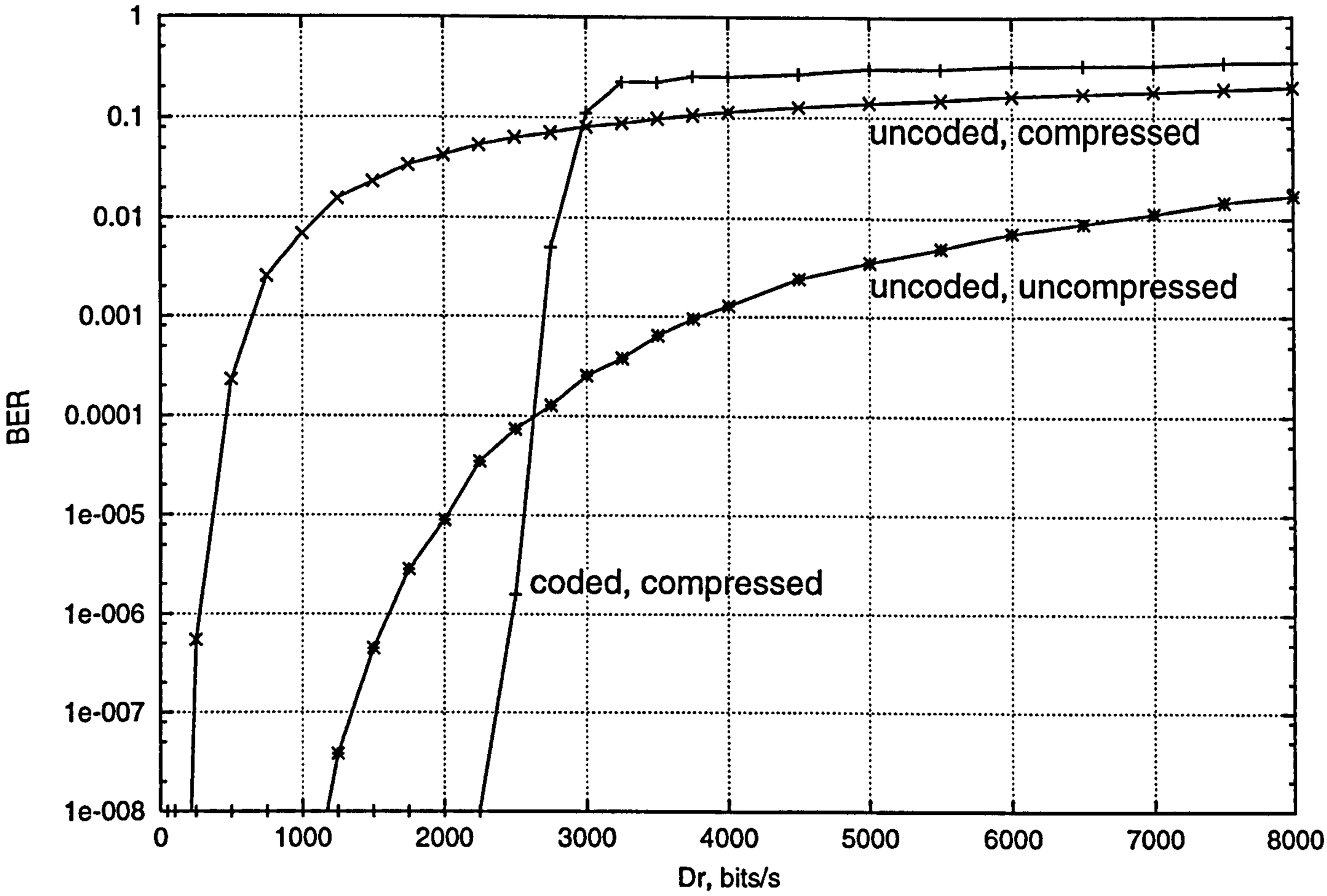

1
0.1
0.01
0.001
0.0001
1e-005
1e-006
1e-007
1e-008
BER
0
1000
2000
3000
4000
5000
6000
7000
8000
Dr, bits/s
uncoded, compressed
uncoded, uncompressed
coded, compressed

Fig 7(b)

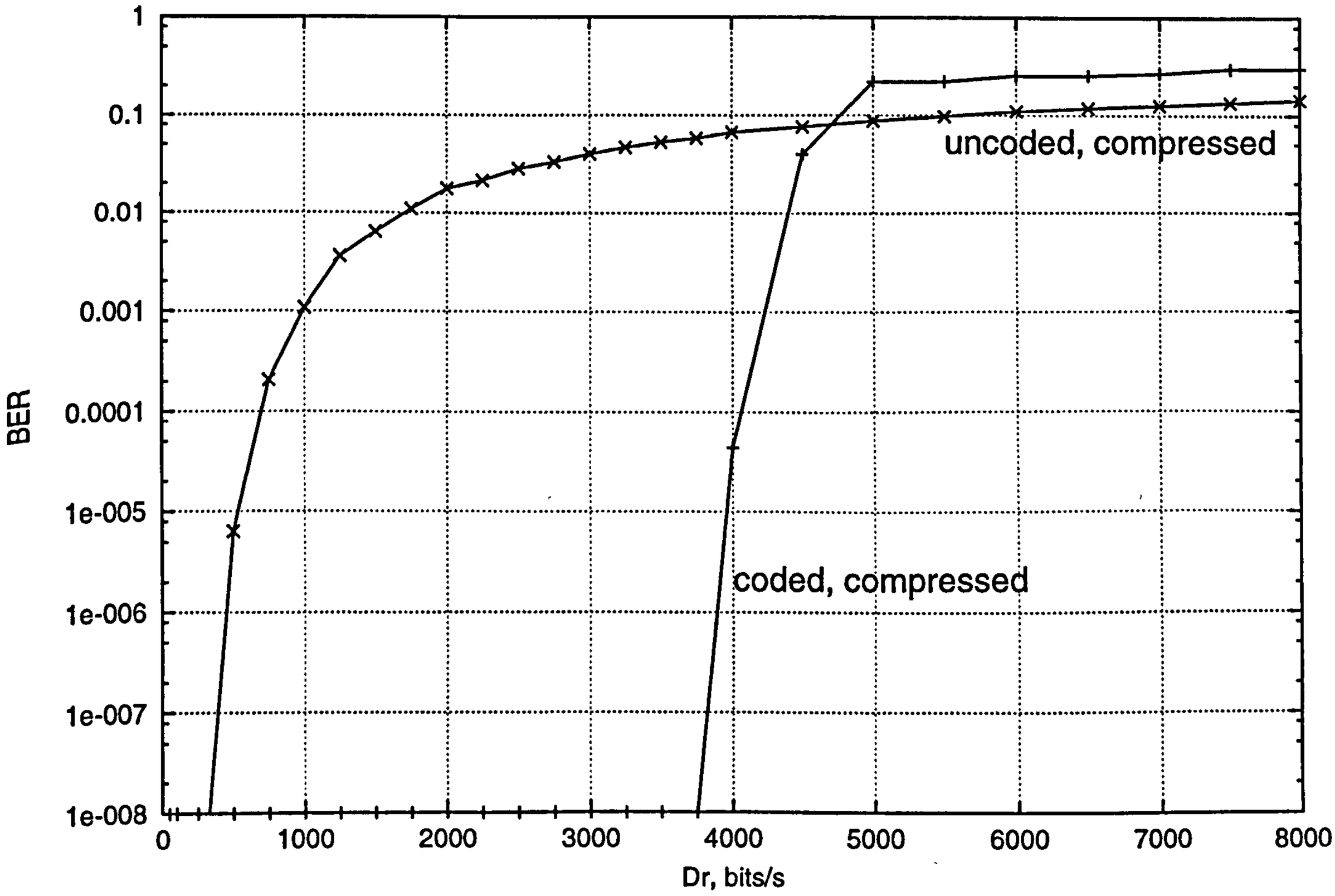
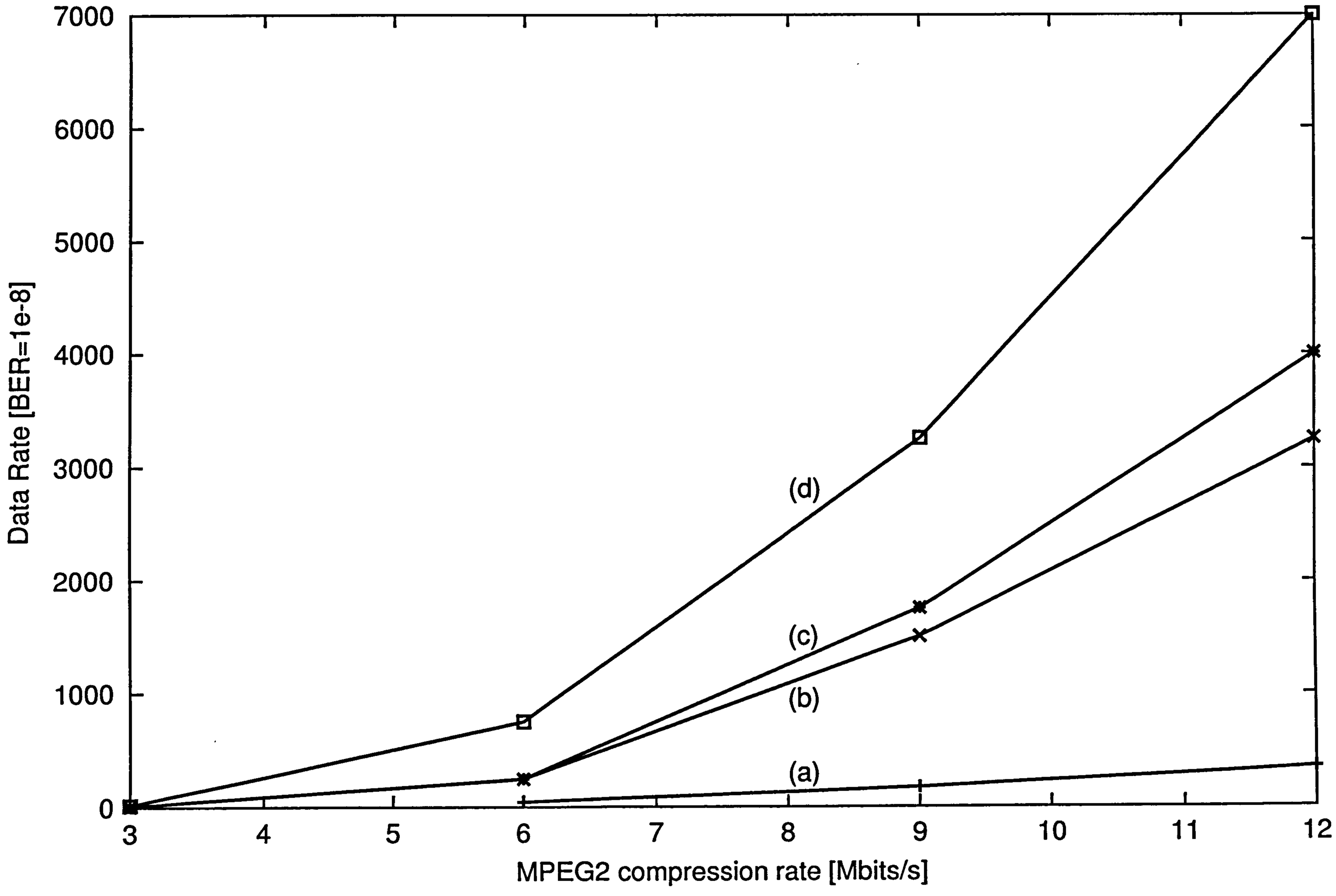

1
0.1
0.01
0.001
0.0001
1e-005
1e-006
1e-007
1e-008
BER
0
1000
2000
3000
4000
5000
6000
7000
8000
Dr, bits/s
uncoded, compressed
coded, compressed

Fig 8


7000
6000
5000
4000
3000
2000
1000
0
Data Rate [BER=1e-8]
3
4
5
6
7
8
9
10
11
12
MPEG2 compression rate [Mbits/s]
(d)
(c)
(b)
(a)

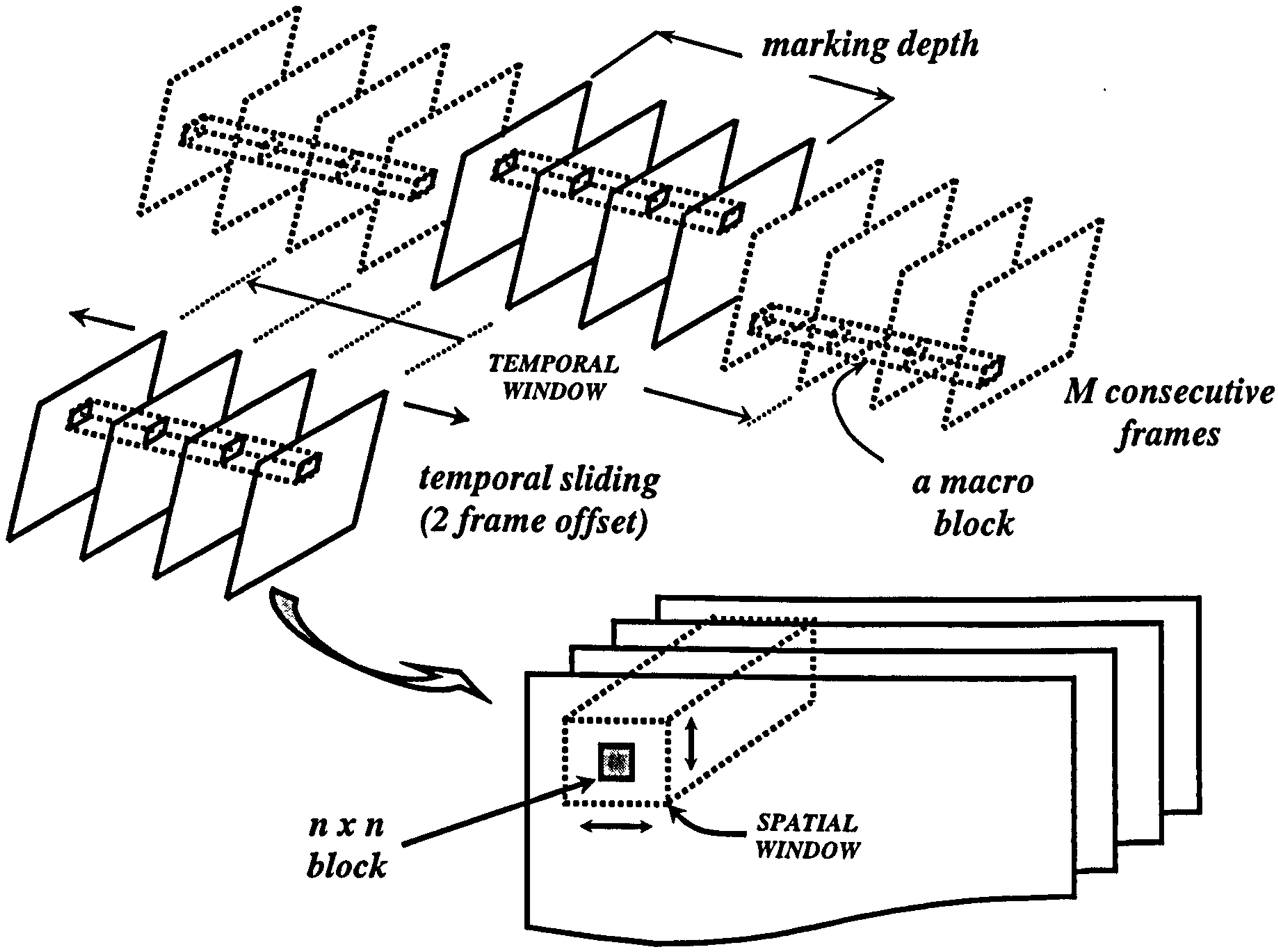
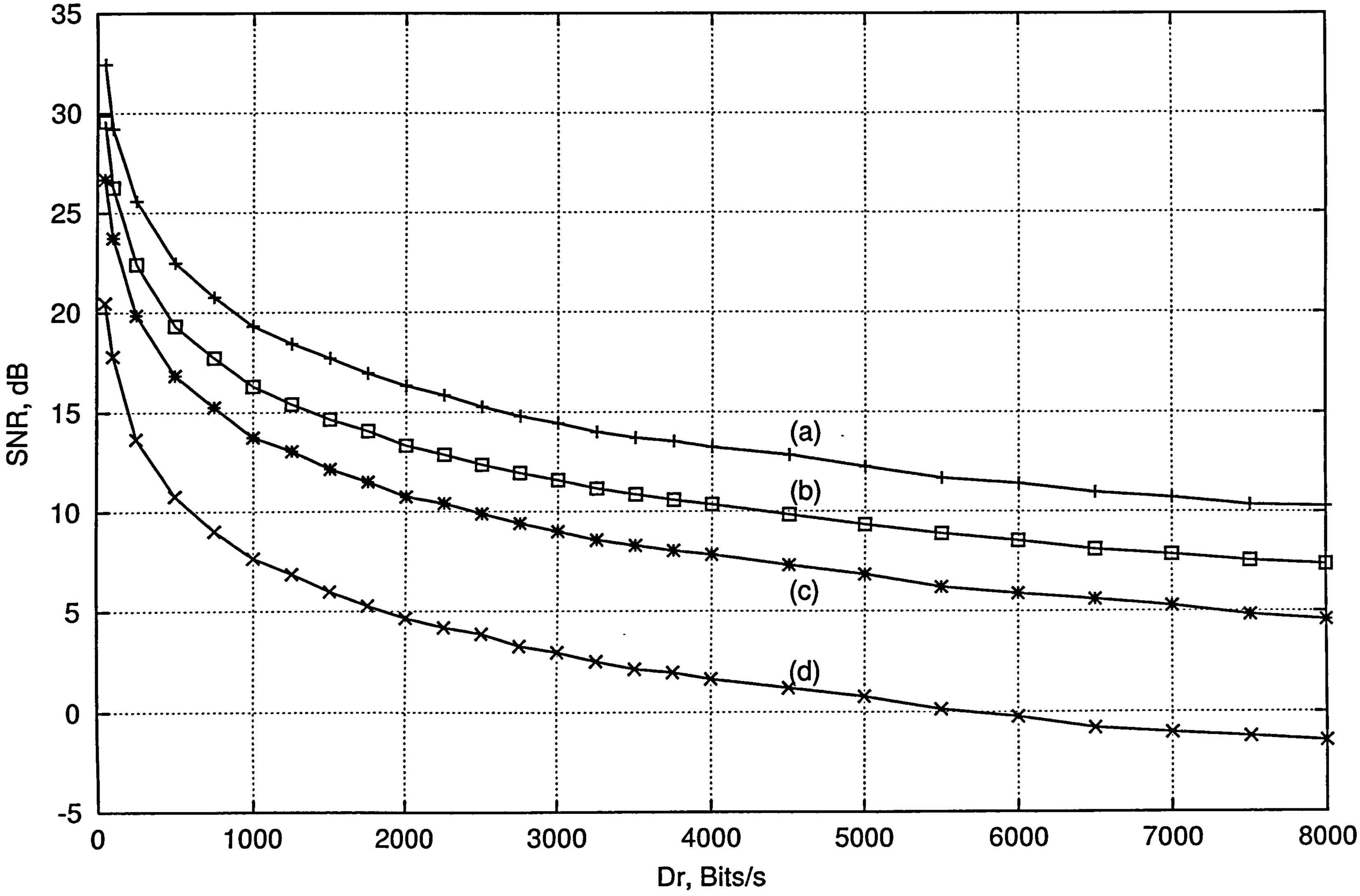
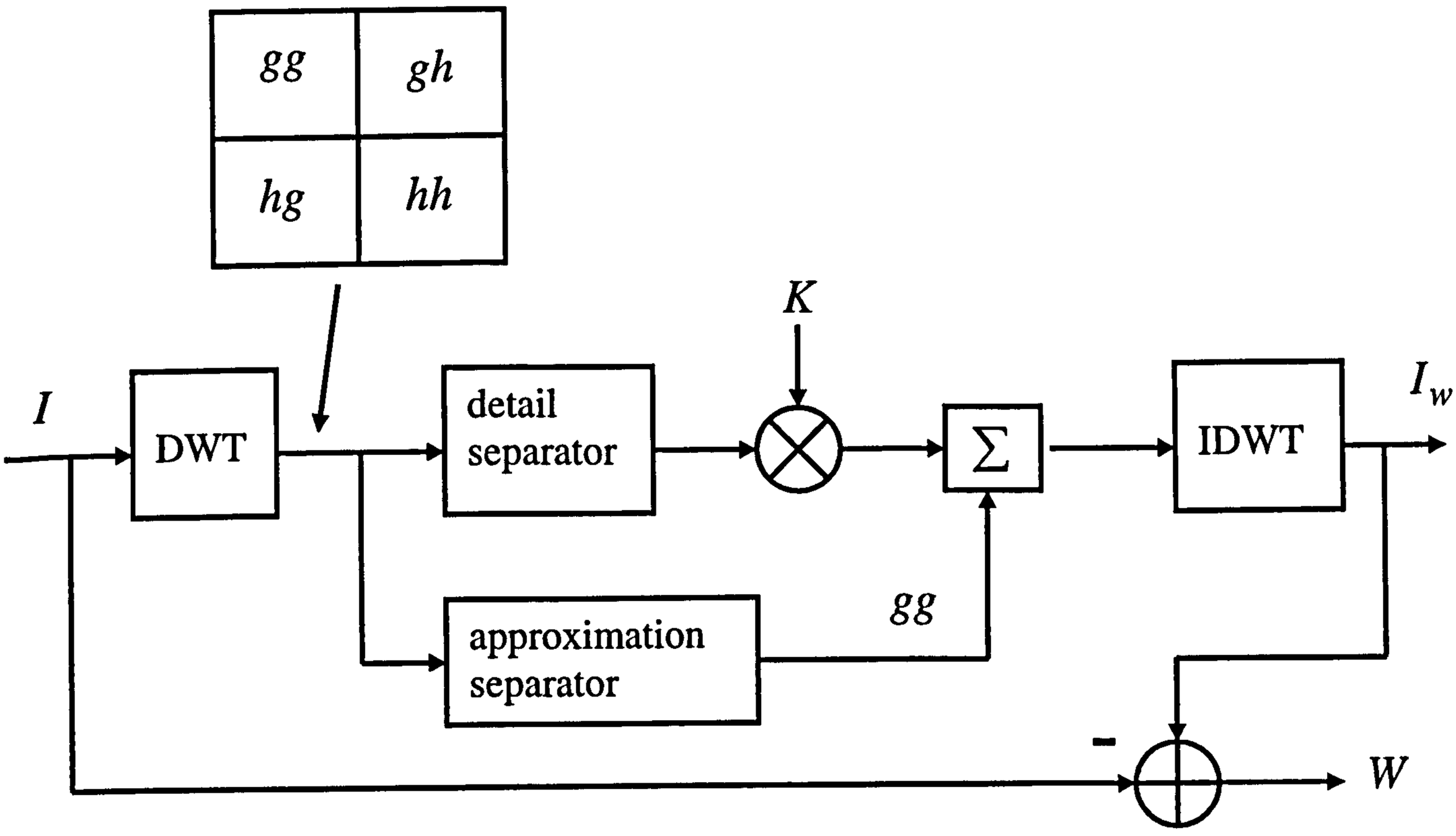


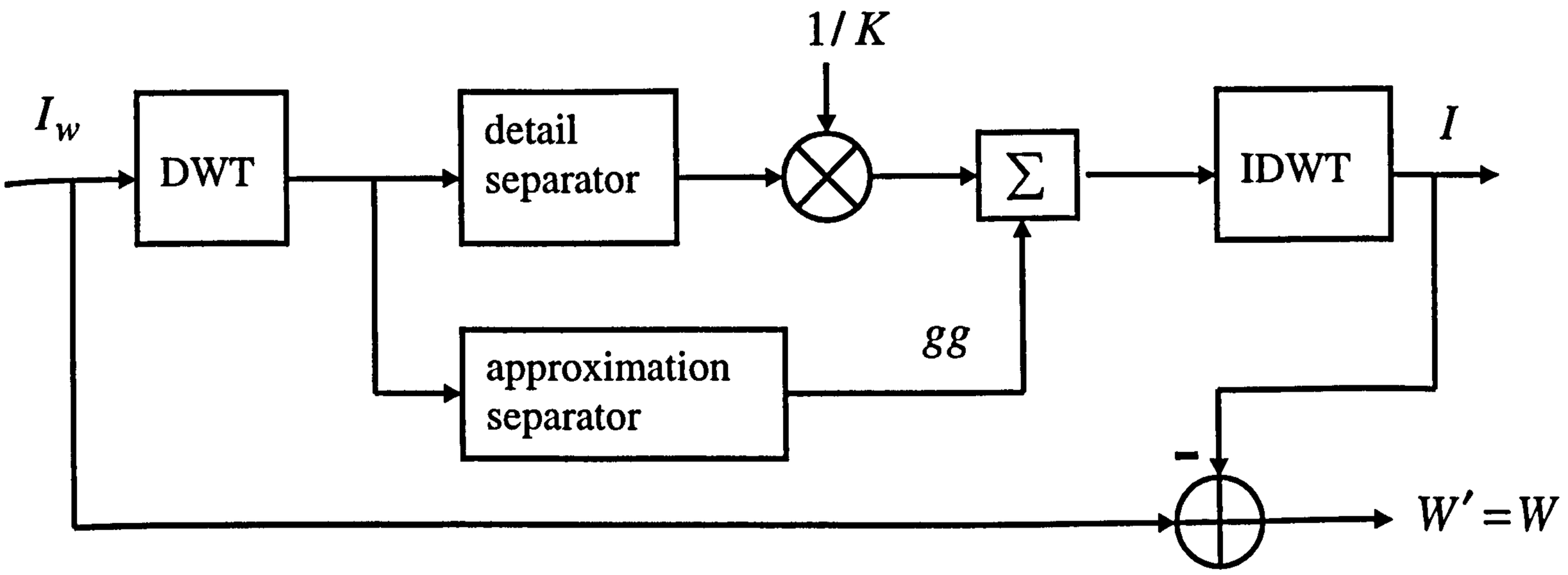
Fig. 9 : 3-D sliding correlator

Fig 10


35
30
25
20
15
10
5
0
-5
SNR, dB
0
1000
2000
3000
4000
5000
6000
7000
8000
Dr, Bits/s
(a)
(b)
(c)
(d)



(a)



(b)

Fig. 11 : DWT domain watermarking

(a) embedding, (b) retrieval, assuming no attack